

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 23, 2005, 12:55:41 ; Search time 88 Seconds
(without alignments)
1644.317 Million cell updates/sec

Title: US-09-421-213-2

Perfect score: 4681

Sequence: 1 MGS DRARKGGGPKDFGAGL.....PGVTRPLRFDWIKENTGV 855

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 781663 seqs, 169239676 residues

Total number of hits satisfying chosen parameters: 781663

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending Patents AA New:*

- 1: /cgn2_6/ptodata/1/paa/PCT_NEW_COMB.pep.*
- 2: /cgn2_6/ptodata/1/paa/US06_NEW_COMB.pep.*
- 3: /cgn2_6/ptodata/1/paa/US07_NEW_COMB.pep.*
- 4: /cgn2_6/ptodata/1/paa/US08_NEW_COMB.pep.*
- 5: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep.*
- 6: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep.*
- 7: /cgn2_6/ptodata/1/paa/US11_NEW_COMB.pep.*
- 8: /cgn2_6/ptodata/1/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4681	100.0	855	6	US-10-612-466B-2
2	4681	100.0	855	7	US-11-104-110-1
3	4681	100.0	855	7	US-11-104-111-22
4	4676	99.9	855	1	PCT-US05-10454-215
5	4676	99.9	855	5	US-09-410-362E-2
6	4676	99.9	855	7	US-11-154-939-650
7	4676	99.9	855	7	US-11-157-575-650
8	4676	99.9	855	8	US-60-687-846-16
9	4676	99.9	855	8	US-60-701-050-901
10	4676	99.9	855	7	US-11-154-939-651
11	4676	99.9	855	7	US-11-167-575-651
12	4676	99.9	855	8	US-60-687-846-15
13	4676	99.9	855	8	US-60-701-050-900
14	4642	99.2	853	6	US-10-530-187-242
15	4319	92.3	932	6	US-10-450-763-51801
16	4175.5	89.2	782	7	US-11-050-926-312
17	1414	30.2	258	7	US-11-104-110-2
18	1414	30.2	258	7	US-11-104-111-23
19	1352	28.9	362	6	US-10-450-763-51800
20	1319	28.2	241	5	US-09-410-362C-2
21	1319	28.2	241	5	US-09-410-362B-3
22	1319	28.2	241	5	US-09-410-362B-35
23	1319	28.2	241	6	US-10-612-466B-4
24	1124	24.0	811	8	US-60-685-372-1264
25	1124	24.0	811	8	US-60-692-282-314

26	1122	24.0	802	5	US-09-978-191C-169	Sequence 169, App
27	1122	24.0	802	7	US-11-129-762-169	Sequence 169, App
28	1122	24.0	802	7	US-11-037-243-113	Sequence 113, App
29	1113.5	23.8	850	7	US-11-037-243-108	Sequence 108, App
30	990.5	21.2	453	6	US-10-450-763-51797	Sequence 51797, A
31	890.5	19.0	152	5	US-09-410-362C-8	Sequence 8, Appli
32	890.5	19.0	152	5	US-09-410-362E-41	Sequence 41, Appl
33	816.5	17.4	152	5	US-09-410-362C-10	Sequence 10, Appl
34	816.5	17.4	152	5	US-09-410-362E-43	Sequence 43, Appl
35	718.5	15.3	1128	7	US-11-037-243-97	Sequence 97, Appl
36	717.5	15.3	531	6	US-10-522-668-2	Sequence 2, Appli
37	692	14.8	1042	1	PCT-US05-06052-1	Sequence 1, Appli
38	692	14.8	1042	6	US-10-543-003-4	Sequence 4, Appli
39	692	14.8	1042	7	US-11-067-811-1	Sequence 1, Appli
40	682.5	14.6	1019	8	US-60-710-726-1836	Sequence 1836, Ap
41	679.5	14.5	1019	1	PCT-US03-10870-2243	Sequence 2243, Ap
42	677	14.5	126	5	US-09-410-362E-22	Sequence 22, Appl
43	663.5	14.2	1113	1	PCT-US05-06052-4	Sequence 4, Appli
44	663.5	14.2	1113	6	US-10-543-003-3	Sequence 3, Appli
45	663.5	14.2	1113	7	US-11-067-811-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1

US-10-612-466B-2

; Sequence 2, Application US/10612466B

; GENERAL INFORMATION: Madison, Edwin

; APPLICANT: Yeh, Juinn-Chern

; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING SERINE PROTEASE 16, THE

; FILE REFERENCE: 24745-1625

; CURRENT APPLICATION NUMBER: US/10/612,466B

; PRIOR FILING DATE: 2003-07-01

; PRIOR APPLICATION NUMBER: 60/394,347

; NUMBER OF SEQ ID NOS: 22

; SOFTWARE: FASTSEQ for Windows Version 4.0

; SEQ ID NO 2

; LENGTH: 855

; TYPE: PRT

; ORGANISM: Homo Sapien

US-10-612-466B-2

Query Match	100.0%;	Score	4681;	DB	6;	Length	855;
Best Local Similarity	100.0%;	Pred. No.	0;				
Matches	855;	Conservative	0;	Mismatches	0;	Indels	0;
						Gaps	0;
QY	1	MGS DRARKGGGPKDFGAGLKYNSRHEKYNGLSEGVFLPVNNVKKVKGHPGRVWVLA	60				
Db	1	MGS DRARKGGGPKDFGAGLKYNSRHEKYNGLSEGVFLPVNNVKKVKGHPGRVWVLA	60				
QY	61	VLIGLLLVLLGIGFLVHLYQYDRVRVQKVFNGYMRITNENFVDAYENSSTFEVSLASKV	120				
Db	61	VLIGLLLVLLGIGFLVHLYQYDRVRVQKVFNGYMRITNENFVDAYENSSTFEVSLASKV	120				
QY	121	KDALKLLYGVFLGPHYHKSATVAFSEGSVYAYWSEFSIQHLVEEAERVMAEERVVM	180				
Db	121	KDALKLLYGVFLGPHYHKSATVAFSEGSVYAYWSEFSIQHLVEEAERVMAEERVVM	180				
QY	181	LPPRARSLKSFVVTTSVAFPTDSKTQRTQDNSCSFGLHARGVELMRFTTTPGPPDPPYPA	240				
Db	181	LPPRARSLKSFVVTTSVAFPTDSKTQRTQDNSCSFGLHARGVELMRFTTTPGPPDPPYPA	240				
QY	241	HARCQWALRGDADSVLSLTFRSFDLASCDESGSDLVTVYNTLSPMPEHALVQLCGTYPPS	300				
Db	241	HARCQWALRGDADSVLSLTFRSFDLASCDESGSDLVTVYNTLSPMPEHALVQLCGTYPPS	300				
QY	301	YNLTHSSQNVLLITLITNTERRRHGFETATFQLPDMSSCGGRLRKAQQTNSPYYPGHY	360				

Db 301 YNLTFHSSQNVLLITLITNTERRHPGFEATFFQLPMSSCGGRLRKAQGTFSPPYPGHY 360
Qy 361 PPNIDCTWNIIEVPNNQHVSKFFYLLLEPGVPAGTCTPKDYVEINGEKYCGERSQFVTS 420
Db 361 PPNIDCTWNIIEVPNNQHVSKFFYLLLEPGVPAGTCTPKDYVEINGEKYCGERSQFVTS 420
Qy 421 NSNKITVRPHSDQSYDTDTGFLAEYLSYDSDPCPGQFTCTGRCIRKELRCDCGWADCTDH 480
Db 421 NSNKITVRPHSDQSYDTDTGFLAEYLSYDSDPCPGQFTCTGRCIRKELRCDCGWADCTDH 480
Qy 481 SDELNCSCDAGHFTCKNKFCKPLFWVCDVNDGDNDEQSCSPAOTFRCSNGKCLSK 540
Db 481 SDELNCSCDAGHFTCKNKFCKPLFWVCDVNDGDNDEQSCSPAOTFRCSNGKCLSK 540
Qy 541 SQQCNKGKDCGSDGSDASCPKVNVTCTKHTYRCLNGLCLSKGNPECDGKEDCSGSD 600
Db 541 SQQCNKGKDCGSDGSDASCPKVNVTCTKHTYRCLNGLCLSKGNPECDGKEDCSGSD 600
Qy 601 DCDCGLRSFTRQARVVGGTDADGEMPMQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
Db 601 DCDCGLRSFTRQARVVGGTDADGEMPMQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
Qy 661 DRGFRYSDDPTQWTAFLGLHDQSORAPGVQERRLRKRIISHPPFNDFTFDYDIALLELEKP 720
Db 661 DRGFRYSDDPTQWTAFLGLHDQSORAPGVQERRLRKRIISHPPFNDFTFDYDIALLELEKP 720
Qy 721 AYSYSSMVRPCLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTCENLL 780
Db 721 AYSYSSMVRPCLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTCENLL 780
Qy 781 PQOITPRMCMVGLSGVDSCQSDGSGPLSSVEADGRIFQAGVSWGDCQAQRNKPQYVT 840
Db 781 PQOITPRMCMVGLSGVDSCQSDGSGPLSSVEADGRIFQAGVSWGDCQAQRNKPQYVT 840
Qy 841 RLPLFRDWIKENTGV 855
Db 841 RLPLFRDWIKENTGV 855

RESULT 2

US-11-104-110-1
; Sequence 1, Application US/11104110
; GENERAL INFORMATION:
; APPLICANT: Ruggles, Sandra
; APPLICANT: Nguyen, Jack
; TITLE OF INVENTION: CLEAVAGE OF VEGF AND VEGF RECEPTOR BY WILDTYPE AND MUTANT MT-SPI
; FILE REFERENCE: 25840-502
; CURRENT APPLICATION NUMBER: US/11/104,110
; PRIOR FILING DATE: 2005-04-12
; PRIOR APPLICATION NUMBER: 60/561,720
; PRIOR FILING DATE: 2004-04-12
; PRIOR APPLICATION NUMBER: 10/677,977
; PRIOR FILING DATE: 2003-10-02
; PRIOR APPLICATION NUMBER: 60/415,388
; PRIOR FILING DATE: 2002-10-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-104-110-1

Query Match 100.0%; Score 4681; DB 7; Length 855;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 855; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MGSDDRKGGGGKDFGAGLKYNRSHEKVGLEEGVEFLPNNVKKVEKHGPGRWVLA 60
Db 1 MGSDDRKGGGGKDFGAGLKYNRSHEKVGLEEGVEFLPNNVKKVEKHGPGRWVLA 60
Qy 61 VLIGLLLVLLGIGFLVHMQYRDVRQKVFNGYMRITNENFVDAVENSNSTEFVSLASKV 120

Db 61 VLIGLLLVLLGIGFLVHMQYRDVRQKVFNGYMRITNENFVDAVENSNSTEFVSLASKV 120
Qy 121 KDALKLISYGVPPFLGPHKESAVTAFSEGSAVIATWSEFSIPOHLVEAEAVMAEERVVM 180
Db 121 KDALKLISYGVPPFLGPHKESAVTAFSEGSAVIATWSEFSIPOHLVEAEAVMAEERVVM 180
Qy 181 LPPRARSLSKSVTVSVVAFPTDSKTIVKTDONSCSFGHLHARGVELMRFTTTPGFDPSPYPA 240
Db 181 LPPRARSLSKSVTVSVVAFPTDSKTIVKTDONSCSFGHLHARGVELMRFTTTPGFDPSPYPA 240
Qy 241 HARCQWALRGDADSVLSLTPRSFDLASCDERGSDLVTVYNTLSPMEPHALVOLCGTYPSP 300
Db 241 HARCQWALRGDADSVLSLTPRSFDLASCDERGSDLVTVYNTLSPMEPHALVOLCGTYPSP 300
Qy 301 YNLTFHSSQNVLLITLITNTERRHPGFEATFFQLPMSSCGGRLRKAQGTFSPPYPGHY 360
Db 301 YNLTFHSSQNVLLITLITNTERRHPGFEATFFQLPMSSCGGRLRKAQGTFSPPYPGHY 360
Qy 361 PPNIDCTWNIIEVPNNQHVSKFFYLLLEPGVPAGTCTPKDYVEINGEKYCGERSQFVTS 420
Db 361 PPNIDCTWNIIEVPNNQHVSKFFYLLLEPGVPAGTCTPKDYVEINGEKYCGERSQFVTS 420
Qy 421 NSNKITVRPHSDQSYDTDTGFLAEYLSYDSDPCPGQFTCTGRCIRKELRCDCGWADCTDH 480
Db 421 NSNKITVRPHSDQSYDTDTGFLAEYLSYDSDPCPGQFTCTGRCIRKELRCDCGWADCTDH 480
Qy 481 SDELNCSCDAGHFTCKNKFCKPLFWVCDVNDGDNDEQSCSPAOTFRCSNGKCLSK 540
Db 481 SDELNCSCDAGHFTCKNKFCKPLFWVCDVNDGDNDEQSCSPAOTFRCSNGKCLSK 540
Qy 541 SQQCNKGKDCGSDGSDASCPKVNVTCTKHTYRCLNGLCLSKGNPECDGKEDCSGSD 600
Db 541 SQQCNKGKDCGSDGSDASCPKVNVTCTKHTYRCLNGLCLSKGNPECDGKEDCSGSD 600
Qy 601 DCDCGLRSFTRQARVVGGTDADGEMPMQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
Db 601 DCDCGLRSFTRQARVVGGTDADGEMPMQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
Qy 661 DRGFRYSDDPTQWTAFLGLHDQSORAPGVQERRLRKRIISHPPFNDFTFDYDIALLELEKP 720
Db 661 DRGFRYSDDPTQWTAFLGLHDQSORAPGVQERRLRKRIISHPPFNDFTFDYDIALLELEKP 720
Qy 721 AYSYSSMVRPCLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTCENLL 780
Db 721 AYSYSSMVRPCLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTCENLL 780
Qy 781 PQOITPRMCMVGLSGVDSCQSDGSGPLSSVEADGRIFQAGVSWGDCQAQRNKPQYVT 840
Db 781 PQOITPRMCMVGLSGVDSCQSDGSGPLSSVEADGRIFQAGVSWGDCQAQRNKPQYVT 840
Qy 841 RLPLFRDWIKENTGV 855
Db 841 RLPLFRDWIKENTGV 855

RESULT 3

US-11-104-111-22
; Sequence 22, Application US/11104111
; GENERAL INFORMATION:
; APPLICANT: Waugh Ruggles, Sandra
; APPLICANT: Nguyen, Jack
; TITLE OF INVENTION: CLEAVAGE OF VEGF AND VEGF RECEPTOR BY WILD-TYPE AND MUTANT
; TITLE OF INVENTION: PROTEASES
; FILE REFERENCE: 25840-503
; CURRENT APPLICATION NUMBER: US/11/104,111
; CURRENT FILING DATE: 2005-04-12
; PRIOR APPLICATION NUMBER: 60/561,671
; PRIOR FILING DATE: 2004-04-12
; PRIOR APPLICATION NUMBER: 10/677,977
; PRIOR FILING DATE: 2003-10-02
; PRIOR APPLICATION NUMBER: 60/415,388
; PRIOR FILING DATE: 2002-10-02
; NUMBER OF SEQ ID NOS: 43

; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 22
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-104-111-22

Query Match 100.0%; Score 4681; DB 7; Length 855;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 855; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGSDRARGGGGKDFGAGLKYNRHEKVNGLGVEFLPNNVKKVEKHGGRWVLA 60
DB 1 MGSDRARGGGGKDFGAGLKYNRHEKVNGLGVEFLPNNVKKVEKHGGRWVLA 60

QY 61 VLIGLLVLLGIGFLVHLYQVDRVQVFNQYMRITNENFVDAYENSNSTFVSLASKV 120
DB 61 VLIGLLVLLGIGFLVHLYQVDRVQVFNQYMRITNENFVDAYENSNSTFVSLASKV 120

QY 121 KDALKLYSGVPFLGPHYKHESAVTAFSEGSVIATYWSSEFSIPQHLVBEAERVMAEERVVM 180
DB 121 KDALKLYSGVPFLGPHYKHESAVTAFSEGSVIATYWSSEFSIPQHLVBEAERVMAEERVVM 180

QY 181 LPPRARSLSKSVVTSVAFPTDSKTQVTDNSCSFGLHARGVELMRFTTGFPPDSPYPA 240
DB 181 LPPRARSLSKSVVTSVAFPTDSKTQVTDNSCSFGLHARGVELMRFTTGFPPDSPYPA 240

QY 241 HARCQWALRGDADSVLSLTFRSFDLASCDERGSGLVTVYNTLSPMEPHALVOLCGTYP 300
DB 241 HARCQWALRGDADSVLSLTFRSFDLASCDERGSGLVTVYNTLSPMEPHALVOLCGTYP 300

QY 301 YNLTFHSSQNVLLITLINTERRHPGFEATFQLPMSSCGRLRKAQGTNSPYPGHY 360
DB 301 YNLTFHSSQNVLLITLINTERRHPGFEATFQLPMSSCGRLRKAQGTNSPYPGHY 360

QY 361 PPNIDCTWNIIEVNNQHVSKFYPFLLEPGVPAGTCKDYVEINGEKYCERSQFVVTS 420
DB 361 PPNIDCTWNIIEVNNQHVSKFYPFLLEPGVPAGTCKDYVEINGEKYCERSQFVVTS 420

QY 421 NSNKITVRFHSDQSYTDGFLAEYLSYDSDPCPGQFTCRGRCIRKELRCDGWADCTDH 480
DB 421 NSNKITVRFHSDQSYTDGFLAEYLSYDSDPCPGQFTCRGRCIRKELRCDGWADCTDH 480

QY 481 SDELNCSCDAGHGFCTCKNFKCKPLFWVCDVNDGDNSEDEGSCPAQTFRCNSGKCLSK 540
DB 481 SDELNCSCDAGHGFCTCKNFKCKPLFWVCDVNDGDNSEDEGSCPAQTFRCNSGKCLSK 540

QY 541 SQQCNKGDDCGDSDSEASCPCNVVVTCTKHYRCLNGLCLSGNPECDGKEDCDSDGDEK 600
DB 541 SQQCNKGDDCGDSDSEASCPCNVVVTCTKHYRCLNGLCLSGNPECDGKEDCDSDGDEK 600

QY 601 DCDGLRSFTROARVVGTDGDEGWPQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
DB 601 DCDGLRSFTROARVVGTDGDEGWPQVSLHALGQGHICGASLISPNWLVSAAHCYID 660

QY 661 DRGFRYSDPTQWTAFLGLHDQSQRSAPGVQERRLKRISHPFFNDFTFDYDIALLELEKP 720
DB 661 DRGFRYSDPTQWTAFLGLHDQSQRSAPGVQERRLKRISHPFFNDFTFDYDIALLELEKP 720

QY 721 AYSMSWRPPICLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTTCE 780
DB 721 AYSMSWRPPICLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTTCE 780

QY 781 PQOITPRMCMVGLSGGVDSCQSGGGLPSSVEADGRIFQAGVSWGDCQARNKPGVYT 840
DB 781 PQOITPRMCMVGLSGGVDSCQSGGGLPSSVEADGRIFQAGVSWGDCQARNKPGVYT 840

QY 841 RLPLFRDWIKENTGV 855
DB 841 RLPLFRDWIKENTGV 855

PCT-US05-10454-215
; Sequence 215, Application PC/TUS0510454
; GENERAL INFORMATION:
; APPLICANT: Ford, Shirin K.
; APPLICANT: Perkins, Nancy-Anne A.
; APPLICANT: Jackson, Donald G.
; TITLE OF INVENTION: BIOMARKERS AND METHODS FOR DETERMINING SENSITIVITY TO EPIDERMAL
; FILE OF INVENTION: GROWTH FACTOR RECEPTOR MODULATORS IN NON-SMALL CELL LUNG CANCER
; FILE REFERENCE: 10219 PCT
; CURRENT APPLICATION NUMBER: PCT/US05/10454
; CURRENT FILING DATE: 2005-04-05
; PRIOR APPLICATION NUMBER: 60/556,903
; PRIOR FILING DATE: 2004-03-26
; NUMBER OF SEQ ID NOS: 294
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 215
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US05-10454-215

Query Match 99.9%; Score 4676; DB 1; Length 855;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 854; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSDRARGGGGKDFGAGLKYNRHEKVNGLGVEFLPNNVKKVEKHGGRWVLA 60
DB 1 MGSDRARGGGGKDFGAGLKYNRHEKVNGLGVEFLPNNVKKVEKHGGRWVLA 60

QY 61 VLIGLLVLLGIGFLVHLYQVDRVQVFNQYMRITNENFVDAYENSNSTFVSLASKV 120
DB 61 VLIGLLVLLGIGFLVHLYQVDRVQVFNQYMRITNENFVDAYENSNSTFVSLASKV 120

QY 121 KDALKLYSGVPFLGPHYKHESAVTAFSEGSVIATYWSSEFSIPQHLVBEAERVMAEERVVM 180
DB 121 KDALKLYSGVPFLGPHYKHESAVTAFSEGSVIATYWSSEFSIPQHLVBEAERVMAEERVVM 180

QY 181 LPPRARSLSKSVVTSVAFPTDSKTQVTDNSCSFGLHARGVELMRFTTGFPPDSPYPA 240
DB 181 LPPRARSLSKSVVTSVAFPTDSKTQVTDNSCSFGLHARGVELMRFTTGFPPDSPYPA 240

QY 241 HARCQWALRGDADSVLSLTFRSFDLASCDERGSGLVTVYNTLSPMEPHALVOLCGTYP 300
DB 241 HARCQWALRGDADSVLSLTFRSFDLASCDERGSGLVTVYNTLSPMEPHALVOLCGTYP 300

QY 301 YNLTFHSSQNVLLITLINTERRHPGFEATFQLPMSSCGRLRKAQGTNSPYPGHY 360
DB 301 YNLTFHSSQNVLLITLINTERRHPGFEATFQLPMSSCGRLRKAQGTNSPYPGHY 360

QY 361 PPNIDCTWNIIEVNNQHVSKFYPFLLEPGVPAGTCKDYVEINGEKYCERSQFVVTS 420
DB 361 PPNIDCTWNIIEVNNQHVSKFYPFLLEPGVPAGTCKDYVEINGEKYCERSQFVVTS 420

QY 421 NSNKITVRFHSDQSYTDGFLAEYLSYDSDPCPGQFTCRGRCIRKELRCDGWADCTDH 480
DB 421 NSNKITVRFHSDQSYTDGFLAEYLSYDSDPCPGQFTCRGRCIRKELRCDGWADCTDH 480

QY 481 SDELNCSCDAGHGFCTCKNFKCKPLFWVCDVNDGDNSEDEGSCPAQTFRCNSGKCLSK 540
DB 481 SDELNCSCDAGHGFCTCKNFKCKPLFWVCDVNDGDNSEDEGSCPAQTFRCNSGKCLSK 540

QY 541 SQQCNKGDDCGDSDSEASCPCNVVVTCTKHYRCLNGLCLSGNPECDGKEDCDSDGDEK 600
DB 541 SQQCNKGDDCGDSDSEASCPCNVVVTCTKHYRCLNGLCLSGNPECDGKEDCDSDGDEK 600

QY 601 DCDGLRSFTROARVVGTDGDEGWPQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
DB 601 DCDGLRSFTROARVVGTDGDEGWPQVSLHALGQGHICGASLISPNWLVSAAHCYID 660

QY 661 DRGFRYSDPTQWTAFLGLHDQSQRSAPGVQERRLKRISHPFFNDFTFDYDIALLELEKP 720
DB 661 DRGFRYSDPTQWTAFLGLHDQSQRSAPGVQERRLKRISHPFFNDFTFDYDIALLELEKP 720

```
QY 721 AYSMMVRPICLPDASHVFPAGKAIWVTGWHTQYGGTGALILQKGEIRVINQTCENLL 780
Db 721 AYSMMVRPICLPDASHVFPAGKAIWVTGWHTQYGGTGALILQKGEIRVINQTCENLL 780
QY 781 PQOITPRMCMVGLSGVDSGCGPLSSVEADGRIFQAGVSWGDCGCAQRNKGYYT 840
Db 781 PQOITPRMCMVGLSGVDSGCGPLSSVEADGRIFQAGVSWGDCGCAQRNKGYYT 840
QY 841 RLPLFRDWIKENTGV 855
Db 841 RLPLFRDWIKENTGV 855

RESULT 5
US-09-410-362E-2
; Sequence 2, Application US/09410362E
; GENERAL INFORMATION:
; APPLICANT: CRAIK, CHARLES S.
; APPLICANT: YAKEUCHI, TOSHIHIKO
; APPLICANT: SCHUMAN, MARC
; TITLE OF INVENTION: MEMBRANE TYPE SERINE PROTEASE 1 (MT-SP1) AND USES THEREOF
; FILE REFERENCE: 28644-701.201
; CURRENT APPLICATION NUMBER: US/09/410,362E
; CURRENT FILING DATE: 1999-09-30
; NUMBER OF SEQ ID NOS: 75
; SOFTWARE: Patent in version 3.3
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-410-362E-2

Query Match 99.9%; Score 4676; DB 5; Length 855;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 854; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSDRARKGGGGPKDFGAGLKYNRSHKVNGLKEGVEFLPVNNVKKVKEKGGPRGWVLA 60
Db 1 MGSDRARKGGGGPKDFGAGLKYNRSHKVNGLKEGVEFLPVNNVKKVKEKGGPRGWVLA 60
QY 61 VLIGLLLVLLGIGFLVHLYQYRDVVRQVFNQYMRITNENFVDAYENSNSFEFVSLASKV 120
Db 61 VLIGLLLVLLGIGFLVHLYQYRDVVRQVFNQYMRITNENFVDAYENSNSFEFVSLASKV 120
QY 121 KDALKLLYSGVPFLGPHYKESAVTAFSEGSAVIAYWSEFSIPQHLVEAEARVMAEERVVM 180
Db 121 KDALKLLYSGVPFLGPHYKESAVTAFSEGSAVIAYWSEFSIPQHLVEAEARVMAEERVVM 180
QY 181 LPPRARSLSKSFVTVSVVAFPTDSKTQVTDNSCSFGLHARGVELMRFTTTPGPDSPYPA 240
Db 181 LPPRARSLSKSFVTVSVVAFPTDSKTQVTDNSCSFGLHARGVELMRFTTTPGPDSPYPA 240
QY 241 HARCOMALRGDADSVLSLTFERSFDLASCDERGSDDLTVYNTLSMPHEHALVOLCGTYPPS 300
Db 241 HARCOMALRGDADSVLSLTFERSFDLASCDERGSDDLTVYNTLSMPHEHALVOLCGTYPPS 300
QY 301 YNLTFHSSQNVLITLITNTERRHPGFEATFFQLPRMSSCGGLRKAQGTNSPYPGHY 360
Db 301 YNLTFHSSQNVLITLITNTERRHPGFEATFFQLPRMSSCGGLRKAQGTNSPYPGHY 360
QY 361 PPNIDCTWNIIEVPNNQHVKVPKFFYLLEPGVPAGTCKPDYVEINGEKYCGERSQFVVT 420
Db 361 PPNIDCTWNIIEVPNNQHVKVPKFFYLLEPGVPAGTCKPDYVEINGEKYCGERSQFVVT 420
QY 421 NSNKITVRFHSDQSYTDGFLAEVLSYDSDPGQFTCRGTCIRKELRCDGWADCTDH 480
Db 421 NSNKITVRFHSDQSYTDGFLAEVLSYDSDPGQFTCRGTCIRKELRCDGWADCTDH 480
QY 481 SDELNCSADAGHQFTCKNFKCPFLFWCDSDVNDGDNDEGQSCPAQTFRCNSGKCLSK 540
Db 481 SDELNCSADAGHQFTCKNFKCPFLFWCDSDVNDGDNDEGQSCPAQTFRCNSGKCLSK 540
QY 541 SQQNGKDDCGDSDGDEASCPKVNVTCTKHTYRCLNGLCLSKGNPECDGKEDCSGSD 600
```

```
Db 541 SQQNGKDDCGDSDGDEASCPKVNVTCTKHTYRCLNGLCLSKGNPECDGKEDCSGSD 600
QY 601 DCDCLGSLSFTRQARVVGTDGEGWPQVSLHALGQGHICGASLISPNLWLSAAHCYID 660
Db 601 DCDCLGSLSFTRQARVVGTDGEGWPQVSLHALGQGHICGASLISPNLWLSAAHCYID 660
QY 661 DRGFRYSDPTQWTAFLGLHDOSQSAFCVQERRLRKRIISHPFFNDFTDYDIALLELEKP 720
Db 661 DRGFRYSDPTQWTAFLGLHDOSQSAFCVQERRLRKRIISHPFFNDFTDYDIALLELEKP 720
QY 721 AYSMMVRPICLPDASHVFPAGKAIWVTGWHTQYGGTGALILQKGEIRVINQTCENLL 780
Db 721 AYSMMVRPICLPDASHVFPAGKAIWVTGWHTQYGGTGALILQKGEIRVINQTCENLL 780
QY 781 PQOITPRMCMVGLSGVDSGCGPLSSVEADGRIFQAGVSWGDCGCAQRNKGYYT 840
Db 781 PQOITPRMCMVGLSGVDSGCGPLSSVEADGRIFQAGVSWGDCGCAQRNKGYYT 840
QY 841 RLPLFRDWIKENTGV 855
Db 841 RLPLFRDWIKENTGV 855

RESULT 6
US-11-154-939-650
; Sequence 650, Application US/11154939
; GENERAL INFORMATION:
; APPLICANT: BIRSE, Charles
; TITLE OF INVENTION: Breast Disease Targets and Uses Thereof
; FILE REFERENCE: CL001529
; CURRENT APPLICATION NUMBER: US/11/154,939
; CURRENT FILING DATE: 2005-06-17
; NUMBER OF SEQ ID NOS: 6081
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 650
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-154-939-650

Query Match 99.9%; Score 4676; DB 7; Length 855;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 854; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSDRARKGGGGPKDFGAGLKYNRSHKVNGLKEGVEFLPVNNVKKVKEKGGPRGWVLA 60
Db 1 MGSDRARKGGGGPKDFGAGLKYNRSHKVNGLKEGVEFLPVNNVKKVKEKGGPRGWVLA 60
QY 61 VLIGLLLVLLGIGFLVHLYQYRDVVRQVFNQYMRITNENFVDAYENSNSFEFVSLASKV 120
Db 61 VLIGLLLVLLGIGFLVHLYQYRDVVRQVFNQYMRITNENFVDAYENSNSFEFVSLASKV 120
QY 121 KDALKLLYSGVPFLGPHYKESAVTAFSEGSAVIAYWSEFSIPQHLVEAEARVMAEERVVM 180
Db 121 KDALKLLYSGVPFLGPHYKESAVTAFSEGSAVIAYWSEFSIPQHLVEAEARVMAEERVVM 180
QY 181 LPPRARSLSKSFVTVSVVAFPTDSKTQVTDNSCSFGLHARGVELMRFTTTPGPDSPYPA 240
Db 181 LPPRARSLSKSFVTVSVVAFPTDSKTQVTDNSCSFGLHARGVELMRFTTTPGPDSPYPA 240
QY 241 HARCOMALRGDADSVLSLTFERSFDLASCDERGSDDLTVYNTLSMPHEHALVOLCGTYPPS 300
Db 241 HARCOMALRGDADSVLSLTFERSFDLASCDERGSDDLTVYNTLSMPHEHALVOLCGTYPPS 300
QY 301 YNLTFHSSQNVLITLITNTERRHPGFEATFFQLPRMSSCGGLRKAQGTNSPYPGHY 360
Db 301 YNLTFHSSQNVLITLITNTERRHPGFEATFFQLPRMSSCGGLRKAQGTNSPYPGHY 360
QY 361 PPNIDCTWNIIEVPNNQHVKVPKFFYLLEPGVPAGTCKPDYVEINGEKYCGERSQFVVT 420
Db 361 PPNIDCTWNIIEVPNNQHVKVPKFFYLLEPGVPAGTCKPDYVEINGEKYCGERSQFVVT 420
```



```
QY 421 NSNKITVRHSDQSYTDTGFLAEYLSYSDSDPCPGQFTCTGTCIRKELRCGWDGADCTDH 480
Db 421 NSNKITVRHSDQSYTDTGFLAEYLSYSDSDPCPGQFTCTGTCIRKELRCGWDGADCTDH 480
QY 481 SDELNCSCDAGHQFTCKNFKCPKPLFWVCDVNDGDNDEQSCSPAQTFRCNSNGKCLSK 540
Db 481 SDELNCSCDAGHQFTCKNFKCPKPLFWVCDVNDGDNDEQSCSPAQTFRCNSNGKCLSK 540
QY 541 SQQNGKDDCGDSDSEASCPKVNVTCTKHYRCLNGLCLSKGNPECDGKEDCDSDGDEK 600
Db 541 SQQNGKDDCGDSDSEASCPKVNVTCTKHYRCLNGLCLSKGNPECDGKEDCDSDGDEK 600
QY 601 DCDGLRSPTROARVVGTTDADGEPWQVSLHALGQGHICGASLISPWLVSAAHCYID 660
Db 601 DCDGLRSPTROARVVGTTDADGEPWQVSLHALGQGHICGASLISPWLVSAAHCYID 660
QY 661 DRGFRYSPTQWTAFLGLHDQORSAPGVQERRKRIISHPPFNDFTDYDIALLELEKP 720
Db 661 DRGFRYSPTQWTAFLGLHDQORSAPGVQERRKRIISHPPFNDFTDYDIALLELEKP 720
QY 721 AYSWVRPCLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTCENLL 780
Db 721 AYSWVRPCLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTCENLL 780
QY 781 PQOITPRMVCVFLSGGVDSQGDSCGGLSSVEADGRIFQAGVSWGDCGCAQRNKPQYIT 840
Db 781 PQOITPRMVCVFLSGGVDSQGDSCGGLSSVEADGRIFQAGVSWGDCGCAQRNKPQYIT 840
QY 841 RLPLFRDWIKENTGV 855
Db 841 RLPLFRDWIKENTGV 855

RESULT 7
US-11-167-575-650
; Sequence 650, Application US/11167575
; GENERAL INFORMATION:
; APPLICANT: BIRSE, Charles
; TITLE OF INVENTION: Breast Disease Targets and Uses Thereof
; FILE REFERENCE: CL001529
; CURRENT APPLICATION NUMBER: US/11/167,575
; CURRENT FILING DATE: 2005-06-28
; NUMBER OF SEQ ID NOS: 6081
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 650
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-167-575-650

Query Match 99.9%; Score 4676; DB 7; Length 855;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 854; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSDRARGGGGKDFGAGLKYNRSHKXVNGLESGVEFLPVNNVKKVKGKGGPRWVLA 60
Db 1 MGSDRARGGGGKDFGAGLKYNRSHKXVNGLESGVEFLPVNNVKKVKGKGGPRWVLA 60
QY 61 VLIGLLVLLGIGFLVHLLQYRDVRVQKVFNGYMRITNENFVDAYENSNSTEFVSLASKV 120
Db 61 VLIGLLVLLGIGFLVHLLQYRDVRVQKVFNGYMRITNENFVDAYENSNSTEFVSLASKV 120
QY 121 KDALKLLYSVPFLGPHYKESAVTAFSEGSVIAYWSEFSIPQHLVBEAERVMABERVVM 180
Db 121 KDALKLLYSVPFLGPHYKESAVTAFSEGSVIAYWSEFSIPQHLVBEAERVMABERVVM 180
QY 181 LPPARSILKSPVTSVVAFPDTSKTQRTQDNCSFGLHARGVELMFTTTPGPDSPYPA 240
Db 181 LPPARSILKSPVTSVVAFPDTSKTQRTQDNCSFGLHARGVELMFTTTPGPDSPYPA 240
QY 241 HARCQWALRGDADSVLSITFRSFDLASCDEGSDLVTVNTLSPMEPHALVQLCGTYP 300
Db 241 HARCQWALRGDADSVLSITFRSFDLASCDEGSDLVTVNTLSPMEPHALVQLCGTYP 300
```

```
QY 301 YNLTFHSSQNVLLITLINTERRHPGFEATFFQPRMSSCGRLRKAQGTNSPYPGHY 360
Db 301 YNLTFHSSQNVLLITLINTERRHPGFEATFFQPRMSSCGRLRKAQGTNSPYPGHY 360
QY 361 PPNIDCTWNIIEVPNNQHVKSFFFYLLRPGVPAGTCKPKDYVEINGEKYCGERSQFVVT 420
Db 361 PPNIDCTWNIIEVPNNQHVKSFFFYLLRPGVPAGTCKPKDYVEINGEKYCGERSQFVVT 420
QY 421 NSNKITVRHSDQSYTDTGFLAEYLSYSDSDPCPGQFTCTGTCIRKELRCGWDGADCTDH 480
Db 421 NSNKITVRHSDQSYTDTGFLAEYLSYSDSDPCPGQFTCTGTCIRKELRCGWDGADCTDH 480
QY 481 SDELNCSCDAGHQFTCKNFKCPKPLFWVCDVNDGDNDEQSCSPAQTFRCNSNGKCLSK 540
Db 481 SDELNCSCDAGHQFTCKNFKCPKPLFWVCDVNDGDNDEQSCSPAQTFRCNSNGKCLSK 540
QY 541 SQQNGKDDCGDSDSEASCPKVNVTCTKHYRCLNGLCLSKGNPECDGKEDCDSDGDEK 600
Db 541 SQQNGKDDCGDSDSEASCPKVNVTCTKHYRCLNGLCLSKGNPECDGKEDCDSDGDEK 600
QY 601 DCDGLRSPTROARVVGTTDADGEPWQVSLHALGQGHICGASLISPWLVSAAHCYID 660
Db 601 DCDGLRSPTROARVVGTTDADGEPWQVSLHALGQGHICGASLISPWLVSAAHCYID 660
QY 661 DRGFRYSPTQWTAFLGLHDQORSAPGVQERRKRIISHPPFNDFTDYDIALLELEKP 720
Db 661 DRGFRYSPTQWTAFLGLHDQORSAPGVQERRKRIISHPPFNDFTDYDIALLELEKP 720
QY 721 AYSWVRPCLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTCENLL 780
Db 721 AYSWVRPCLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTCENLL 780
QY 781 PQOITPRMVCVFLSGGVDSQGDSCGGLSSVEADGRIFQAGVSWGDCGCAQRNKPQYIT 840
Db 781 PQOITPRMVCVFLSGGVDSQGDSCGGLSSVEADGRIFQAGVSWGDCGCAQRNKPQYIT 840
QY 841 RLPLFRDWIKENTGV 855
Db 841 RLPLFRDWIKENTGV 855

RESULT 8
US-60-687-846-16
; Sequence 16, Application US/60687846
; GENERAL INFORMATION:
; APPLICANT: JOSELOFF, Elizabeth et al.
; TITLE OF INVENTION: STOMACH DISEASE TARGETS AND USES THEREOF
; FILE REFERENCE: CL001606
; CURRENT APPLICATION NUMBER: US/60/687,846
; CURRENT FILING DATE: 2005-06-07
; NUMBER OF SEQ ID NOS: 1565
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-687-846-16

Query Match 99.9%; Score 4676; DB 8; Length 855;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 854; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSDRARGGGGKDFGAGLKYNRSHKXVNGLESGVEFLPVNNVKKVKGKGGPRWVLA 60
Db 1 MGSDRARGGGGKDFGAGLKYNRSHKXVNGLESGVEFLPVNNVKKVKGKGGPRWVLA 60
QY 61 VLIGLLVLLGIGFLVHLLQYRDVRVQKVFNGYMRITNENFVDAYENSNSTEFVSLASKV 120
Db 61 VLIGLLVLLGIGFLVHLLQYRDVRVQKVFNGYMRITNENFVDAYENSNSTEFVSLASKV 120
QY 121 KDALKLLYSVPFLGPHYKESAVTAFSEGSVIAYWSEFSIPQHLVBEAERVMABERVVM 180
Db 121 KDALKLLYSVPFLGPHYKESAVTAFSEGSVIAYWSEFSIPQHLVBEAERVMABERVVM 180
```

Db 121 KDALKLLYSGVPLGPHYKESAVTAFSEGSAVIAIYWSBFSPHQLVVEAERVMABERVVM 180
Qy 181 LPPRARSLSKSVVTSVVAFTDSKTQRTQDNSCSFGLHARGVELMRFTTFCGPPSPYPA 240
Db 181 LPPRARSLSKSVVTSVVAFTDSKTQRTQDNSCSFGLHARGVELMRFTTFCGPPSPYPA 240
Qy 241 HARCQWALRGDADSVLSLTFRFDLASCDERGSDLVTVNTLSPMEPHALVOLCGTYPPS 300
Db 241 HARCQWALRGDADSVLSLTFRFDLASCDERGSDLVTVNTLSPMEPHALVOLCGTYPPS 300
Qy 301 YNLTFHSSQNVLLITLITNTERRHPGFEATFQQLPRMSSCGRLRKAQGTENSPPYPGHY 360
Db 301 YNLTFHSSQNVLLITLITNTERRHPGFEATFQQLPRMSSCGRLRKAQGTENSPPYPGHY 360
Qy 361 PNIDCTWNIIEVNNQHVSKVFFYLLPEPGVAGTCTPKDYVEINGEKYCGERSQFVVT 420
Db 361 PNIDCTWNIIEVNNQHVSKVFFYLLPEPGVAGTCTPKDYVEINGEKYCGERSQFVVT 420
Qy 421 NSNKITVRFHSDQSYDTDTGFLAEYLSYDSSDPCPGQFTCTGRCIRKELRCDGWADCTDH 480
Db 421 NSNKITVRFHSDQSYDTDTGFLAEYLSYDSSDPCPGQFTCTGRCIRKELRCDGWADCTDH 480
Qy 481 SDELNCSCDAGHOFCKNFKCKPLFWCDSVNDGNSDEQSCSPAQOTFRCSNGKCLSK 540
Db 481 SDELNCSCDAGHOFCKNFKCKPLFWCDSVNDGNSDEQSCSPAQOTFRCSNGKCLSK 540
Qy 541 SOQNGKDDCGSDSDASCPKVNVTCTKHTYRCLNGLCLSKGNPECDGKEDCSDGDEK 600
Db 541 SOQNGKDDCGSDSDASCPKVNVTCTKHTYRCLNGLCLSKGNPECDGKEDCSDGDEK 600
Qy 601 DDCGLRSFTRQARVVGTTDAGEWPQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
Db 601 DDCGLRSFTRQARVVGTTDAGEWPQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
Qy 661 DRGFYSDPTQWTAFLGLHDQSORAPGVQERRLKRISHPFFNDFTFDYDIALLELEKP 720
Db 661 DRGFYSDPTQWTAFLGLHDQSORAPGVQERRLKRISHPFFNDFTFDYDIALLELEKP 720
Qy 721 ABEYSSMVRPCLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTTCE 780
Db 721 ABEYSSMVRPCLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTTCE 780
Qy 781 POQITPRMCMVGLSGVSDSCGDSGGLSSVEADGRIFQAGVSVWGDCGCAQRNKP 840
Db 781 POQITPRMCMVGLSGVSDSCGDSGGLSSVEADGRIFQAGVSVWGDCGCAQRNKP 840
Qy 841 RLPLFRDWIKENTGV 855
Db 841 RLPLFRDWIKENTGV 855

RESULT 9

US-60-701-050-901
; Sequence 901, Application US/60701050
; GENERAL INFORMATION:
; APPLICANT: JOSELOFF, Elizabeth et al.
; TITLE OF INVENTION: STOMACH DISEASE TARGETS AND USES THEREOF
; FILE REFERENCE: CLO01621
; CURRENT APPLICATION NUMBER: US/60/701,050
; CURRENT FILING DATE: 2005-07-21
; NUMBER OF SEQ ID NOS: 2554
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 901
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-701-050-901

Query Match 99.9%; Score 4676; DB 8; Length 855;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 854; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MGSDRARKGGGPKDFGAGLKYNSRHEKVNGLSEGVFLPVNNVKVKEKHGPRWVLA 60

Db 1 MGSDRARKGGGPKDFGAGLKYNSRHEKVNGLSEGVFLPVNNVKVKEKHGPRWVLA 60
Qy 61 VLI GLLLVLLGIGLFWHLQYRDVRQVQKFNQYMRITNENFVDAYENSNSTEFVSLASKV 120
Db 61 VLI GLLLVLLGIGLFWHLQYRDVRQVQKFNQYMRITNENFVDAYENSNSTEFVSLASKV 120
Qy 121 KDALKLYSGVPLGPHYKESAVTAFSEGSAVIAIYWSBFSPHQLVVEAERVMABERVVM 180
Db 121 KDALKLYSGVPLGPHYKESAVTAFSEGSAVIAIYWSBFSPHQLVVEAERVMABERVVM 180
Qy 181 LPPRARSLSKSVVTSVVAFTDSKTQRTQDNSCSFGLHARGVELMRFTTFCGPPSPYPA 240
Db 181 LPPRARSLSKSVVTSVVAFTDSKTQRTQDNSCSFGLHARGVELMRFTTFCGPPSPYPA 240
Qy 241 HARCQWALRGDADSVLSLTFRFDLASCDERGSDLVTVNTLSPMEPHALVOLCGTYPPS 300
Db 241 HARCQWALRGDADSVLSLTFRFDLASCDERGSDLVTVNTLSPMEPHALVOLCGTYPPS 300
Qy 301 YNLTFHSSQNVLLITLITNTERRHPGFEATFQQLPRMSSCGRLRKAQGTENSPPYPGHY 360
Db 301 YNLTFHSSQNVLLITLITNTERRHPGFEATFQQLPRMSSCGRLRKAQGTENSPPYPGHY 360
Qy 361 PNIDCTWNIIEVNNQHVSKVFFYLLPEPGVAGTCTPKDYVEINGEKYCGERSQFVVT 420
Db 361 PNIDCTWNIIEVNNQHVSKVFFYLLPEPGVAGTCTPKDYVEINGEKYCGERSQFVVT 420
Qy 421 NSNKITVRFHSDQSYDTDTGFLAEYLSYDSSDPCPGQFTCTGRCIRKELRCDGWADCTDH 480
Db 421 NSNKITVRFHSDQSYDTDTGFLAEYLSYDSSDPCPGQFTCTGRCIRKELRCDGWADCTDH 480
Qy 481 SDELNCSCDAGHOFCKNFKCKPLFWCDSVNDGNSDEQSCSPAQOTFRCSNGKCLSK 540
Db 481 SDELNCSCDAGHOFCKNFKCKPLFWCDSVNDGNSDEQSCSPAQOTFRCSNGKCLSK 540
Qy 541 SOQNGKDDCGSDSDASCPKVNVTCTKHTYRCLNGLCLSKGNPECDGKEDCSDGDEK 600
Db 541 SOQNGKDDCGSDSDASCPKVNVTCTKHTYRCLNGLCLSKGNPECDGKEDCSDGDEK 600
Qy 601 DDCGLRSFTRQARVVGTTDAGEWPQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
Db 601 DDCGLRSFTRQARVVGTTDAGEWPQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
Qy 661 DRGFYSDPTQWTAFLGLHDQSORAPGVQERRLKRISHPFFNDFTFDYDIALLELEKP 720
Db 661 DRGFYSDPTQWTAFLGLHDQSORAPGVQERRLKRISHPFFNDFTFDYDIALLELEKP 720
Qy 721 ABEYSSMVRPCLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTTCE 780
Db 721 ABEYSSMVRPCLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTTCE 780
Qy 781 POQITPRMCMVGLSGVSDSCGDSGGLSSVEADGRIFQAGVSVWGDCGCAQRNKP 840
Db 781 POQITPRMCMVGLSGVSDSCGDSGGLSSVEADGRIFQAGVSVWGDCGCAQRNKP 840
Qy 841 RLPLFRDWIKENTGV 855
Db 841 RLPLFRDWIKENTGV 855

RESULT 10

US-11-154-939-651
; Sequence 651, Application US/11154939
; GENERAL INFORMATION:
; APPLICANT: BIRSE, Charles
; TITLE OF INVENTION: Breast Disease Targets and Uses Thereof
; FILE REFERENCE: CLO01529
; CURRENT APPLICATION NUMBER: US/11/154,939
; CURRENT FILING DATE: 2005-06-17
; NUMBER OF SEQ ID NOS: 6081
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 651
; LENGTH: 856

[illegible]

```

; APPLICANT: BIRSE, Charles
; TITLE OF INVENTION: Breast Disease Targets and Uses Thereof
; FILE REFERENCE: CL001529
; CURRENT APPLICATION NUMBER: US/11/167,575
; CURRENT FILING DATE: 2005-06-28
; NUMBER OF SEQ ID NOS: 6081
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 651
; LENGTH: 856
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-11-167-575-651

```

RESULT 11
US-11-167-575-651
; Sequence 651, Application US/11167575
; GENERAL INFORMATION:

```
QY 841 RLPLFRDWIKENTGV 855
Db 841 RLPLFRDWIKENTGV 855

RESULT 12
US-60-687-846-15
; Sequence 15, Application US/60687846
; GENERAL INFORMATION:
; APPLICANT: JOSELOFF, Elizabeth et al.
; TITLE OF INVENTION: STOMACH DISEASE TARGETS AND USES THEREOF
; FILE REFERENCE: CL001606
; CURRENT APPLICATION NUMBER: US/60/687,846
; CURRENT FILING DATE: 2005-06-07
; NUMBER OF SEQ ID NOS: 1565
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 856
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-687-846-15

Query Match 99.9%; Score 4676; DB 8; Length 856;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 854; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGS DRARKGGGPKDFGAGLKYNRHEKVNGLKEEGVEFLPVNNVKKVKGHGPGRWVLLAA 60
Db 1 MGS DRARKGGGPKDFGAGLKYNRHEKVNGLKEEGVEFLPVNNVKKVKGHGPGRWVLLAA 60

QY 61 VLIGLLLVLLGIGFLVHQLQYRDVRVQKVFNGYMRITNENFVDAYENSNSSTEFVSLASKV 120
Db 61 VLIGLLLVLLGIGFLVHQLQYRDVRVQKVFNGYMRITNENFVDAYENSNSSTEFVSLASKV 120

QY 121 KQALKLYSGVPLGYPHKESAVTAFSEGSVIAYWSEFSIPOHLVEAEARVMAEERVVM 180
Db 121 KQALKLYSGVPLGYPHKESAVTAFSEGSVIAYWSEFSIPOHLVEAEARVMAEERVVM 180

QY 181 LPPRARSLSKSFVVTSVVAFPTDSKTQVTDNSCSFGLHARGVELMRFTTGFPPDSPYPA 240
Db 181 LPPRARSLSKSFVVTSVVAFPTDSKTQVTDNSCSFGLHARGVELMRFTTGFPPDSPYPA 240

QY 241 HARCQWALRGDADSVLSLTFRSPDLASCDERGSDLVTVTNTLSPMEPHALVOLCGTYPSP 300
Db 241 HARCQWALRGDADSVLSLTFRSPDLASCDERGSDLVTVTNTLSPMEPHALVOLCGTYPSP 300

QY 301 YNLTFHSSQNVLITLITNTERRHPGFEATFFQLPRMSSCGRLRKAQGTFSNPPYPGHY 360
Db 301 YNLTFHSSQNVLITLITNTERRHPGFEATFFQLPRMSSCGRLRKAQGTFSNPPYPGHY 360

QY 361 PPNICTWNIIEVPNNQHVSKFVYLLPEPGVPAGTCKPKDYVEINGEKYCGERSQFVVT 420
Db 361 PPNICTWNIIEVPNNQHVSKFVYLLPEPGVPAGTCKPKDYVEINGEKYCGERSQFVVT 420

QY 421 NSNKITVRPHSDQSYTDTGFLAEYLSYDSSDPGQFTCRTRGCRIRKELRCDGWADCTDH 480
Db 421 NSNKITVRPHSDQSYTDTGFLAEYLSYDSSDPGQFTCRTRGCRIRKELRCDGWADCTDH 480

QY 481 SDELNCSCDAGHQFTCKNFKCKPLFWVCDNVNDGNSDEQGCSCPAQTCFRCSNGKCLSK 540
Db 481 SDELNCSCDAGHQFTCKNFKCKPLFWVCDNVNDGNSDEQGCSCPAQTCFRCSNGKCLSK 540

QY 541 SQQCKGKDCGSDGSEASCPKVNVTCTKHTYRCLNGLCLSKGNPECDGKEDCSGSDK 600
Db 541 SQQCKGKDCGSDGSEASCPKVNVTCTKHTYRCLNGLCLSKGNPECDGKEDCSGSDK 600

QY 601 DCDCLGRSPTRQARVVGGTADAGEWPQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
Db 601 DCDCLGRSPTRQARVVGGTADAGEWPQVSLHALGQGHICGASLISPNWLVSAAHCYID 660

QY 661 DRGFRYSDPTQWTAFLGLHDQORSGAPGVQERRLKRIISHPPFNDFTDYDIALLELEKP 720
Db 661 DRGFRYSDPTQWTAFLGLHDQORSGAPGVQERRLKRIISHPPFNDFTDYDIALLELEKP 720
```

```
QY 721 AYS SSVRPICLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTTCE NLL 780
Db 721 AYS SSVRPICLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTTCE NLL 780

QY 781 PQQITPRMVCVGFSLGGVDSQCGSGPLSSVEADGRIFQAGVVSWSGDCGAQRNPGVYT 840
Db 781 PQQITPRMVCVGFSLGGVDSQCGSGPLSSVEADGRIFQAGVVSWSGDCGAQRNPGVYT 840

QY 841 RLPLFRDWIKENTGV 855
Db 841 RLPLFRDWIKENTGV 855

RESULT 13
US-60-701-050-900
; Sequence 900, Application US/60701050
; GENERAL INFORMATION:
; APPLICANT: JOSELOFF, Elizabeth et al.
; TITLE OF INVENTION: STOMACH DISEASE TARGETS AND USES THEREOF
; FILE REFERENCE: CL001621
; CURRENT APPLICATION NUMBER: US/60/701,050
; CURRENT FILING DATE: 2005-07-21
; NUMBER OF SEQ ID NOS: 2554
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 900
; LENGTH: 856
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-701-050-900

Query Match 99.9%; Score 4676; DB 8; Length 856;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 854; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGS DRARKGGGPKDFGAGLKYNRHEKVNGLKEEGVEFLPVNNVKKVKGHGPGRWVLLAA 60
Db 1 MGS DRARKGGGPKDFGAGLKYNRHEKVNGLKEEGVEFLPVNNVKKVKGHGPGRWVLLAA 60

QY 61 VLIGLLLVLLGIGFLVHQLQYRDVRVQKVFNGYMRITNENFVDAYENSNSSTEFVSLASKV 120
Db 61 VLIGLLLVLLGIGFLVHQLQYRDVRVQKVFNGYMRITNENFVDAYENSNSSTEFVSLASKV 120

QY 121 KQALKLYSGVPLGYPHKESAVTAFSEGSVIAYWSEFSIPOHLVEAEARVMAEERVVM 180
Db 121 KQALKLYSGVPLGYPHKESAVTAFSEGSVIAYWSEFSIPOHLVEAEARVMAEERVVM 180

QY 181 LPPRARSLSKSFVVTSVVAFPTDSKTQVTDNSCSFGLHARGVELMRFTTGFPPDSPYPA 240
Db 181 LPPRARSLSKSFVVTSVVAFPTDSKTQVTDNSCSFGLHARGVELMRFTTGFPPDSPYPA 240

QY 241 HARCQWALRGDADSVLSLTFRSPDLASCDERGSDLVTVTNTLSPMEPHALVOLCGTYPSP 300
Db 241 HARCQWALRGDADSVLSLTFRSPDLASCDERGSDLVTVTNTLSPMEPHALVOLCGTYPSP 300

QY 301 YNLTFHSSQNVLITLITNTERRHPGFEATFFQLPRMSSCGRLRKAQGTFSNPPYPGHY 360
Db 301 YNLTFHSSQNVLITLITNTERRHPGFEATFFQLPRMSSCGRLRKAQGTFSNPPYPGHY 360

QY 361 PPNICTWNIIEVPNNQHVSKFVYLLPEPGVPAGTCKPKDYVEINGEKYCGERSQFVVT 420
Db 361 PPNICTWNIIEVPNNQHVSKFVYLLPEPGVPAGTCKPKDYVEINGEKYCGERSQFVVT 420

QY 421 NSNKITVRPHSDQSYTDTGFLAEYLSYDSSDPGQFTCRTRGCRIRKELRCDGWADCTDH 480
Db 421 NSNKITVRPHSDQSYTDTGFLAEYLSYDSSDPGQFTCRTRGCRIRKELRCDGWADCTDH 480

QY 481 SDELNCSCDAGHQFTCKNFKCKPLFWVCDNVNDGNSDEQGCSCPAQTCFRCSNGKCLSK 540
Db 481 SDELNCSCDAGHQFTCKNFKCKPLFWVCDNVNDGNSDEQGCSCPAQTCFRCSNGKCLSK 540

QY 541 SQQCKGKDCGSDGSEASCPKVNVTCTKHTYRCLNGLCLSKGNPECDGKEDCSGSDK 600
Db 541 SQQCKGKDCGSDGSEASCPKVNVTCTKHTYRCLNGLCLSKGNPECDGKEDCSGSDK 600
```

Db 541 SQCCNGKDCGDSDEASCPKVNVTCTKHYRCLNGCLSKGNPECDKEDCSGDSBK 600
Qy 601 DDCGCLRSFTRQARVVGTTDADEGEWPMQVSLHALGQGHICGASLISPNMLVSAACHYD 660
Db 601 DDCGCLRSFTRQARVVGTTDADEGEWPMQVSLHALGQGHICGASLISPNMLVSAACHYD 660
Qy 661 DRGFRYSDPTQMTAFGLHLDQSQSAPGVOERRLKRIISHPPFNFTFDYDIALLELEKP 720
Db 661 DRGFRYSDPTQMTAFGLHLDQSQSAPGVOERRLKRIISHPPFNFTFDYDIALLELEKP 720
Qy 721 AYSVMVRPCLPDASHVFPAGKAIWVTGWGHTQVGGTGCALILQKEIRVINQTCENLL 780
Db 721 AYSVMVRPCLPDASHVFPAGKAIWVTGWGHTQVGGTGCALILQKEIRVINQTCENLL 780
Qy 781 PQQITPRMCMVGFSLGGVDSQCGDGGPLSSVEADGRIFQAGVSWGDCGCAQRNKPQVYT 840
Db 781 PQQITPRMCMVGFSLGGVDSQCGDGGPLSSVEADGRIFQAGVSWGDCGCAQRNKPQVYT 840
Qy 841 RLPLFRDWIKENTGV 855
Db 841 RLPLFRDWIKENTGV 855

RESULT 14
US-10-530-187-242
; Sequence 242, Application US/10530187
; GENERAL INFORMATION:
; APPLICANT: THE, Bin Tean
; APPLICANT: TAKAHASHI, Masayuki
; TITLE OF INVENTION: Molecular Subclassification of Kidney Tumors and the Discovery of
; FILE REFERENCE: 28927-0014
; CURRENT APPLICATION NUMBER: US/10/530,187
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: PCT/US2003/031476
; PRIOR FILING DATE: 2003-10-06
; PRIOR APPLICATION NUMBER: US 60/415,775
; PRIOR FILING DATE: 2002-10-04
; NUMBER OF SEQ ID NOS: 296
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 242
; LENGTH: 853
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-530-187-242

Query Match 99.2%; Score 4642; DB 6; Length 853;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 852; Conservative 0; Mismatches 1; Indels 2; Gaps 2;

Qy 1 MGSDEARKGGGKDPFGAGLKYNSRHEKVGLEEGVEFLPVNNVKKVEKHGPGRWVLLA 60
Db 1 MGSDEARKGGGKDPFGAGLKYNSRHEKVGLEEGVEFLPVNNVKKVEKHGPGRWVLLA 60
Qy 61 VLIGLLLVLLGLGFLVHLQYDVRVQVFNQYMRITNFENFDAYENSNSTEFVSLASKV 120
Db 61 VLIGLLLVLLGLGFLVHLQYDVRVQVFNQYMRITNFENFDAYENSNSTEFVSLASKV 120
Qy 121 KDALKLLYGVVPFLGPYHKESAVTAFSGSVIAYYWSFSPQHLVEBAERVMAEERVVM 180
Db 121 KDALKLLYGVVPFLGPYHKESAVTAFSGSVIAYYWSFSPQHLVEBAERVMAEERVVM 180
Qy 181 LPPRARSLSKSVVTSVVAFFTDSKTQVQTDNSCSFGLHARGVELMRTTTFGPDSPYPA 240
Db 181 LPPRARSLSKSVVTSVVAFFTDSKTQVQTDNSCSFGLHARGVELMRTTTFGPDSPYPA 239
Qy 241 HARCOWALRGDADSVLSITFTFSFDLASCDERGSDDLVTYNTLSMPEHALVOLGCTYPPS 300
Db 240 HA-COWALRGDADSVLSITFTFSFDLASCDERGSDDLVTYNTLSMPEHALVOLGCTYPPS 298
Qy 301 YNLTFHSSQVLLITLINTERRHPGFATFFQLPRMSSCGRLKAKQGTNSPYPGHY 360
Db 299 YNLTFHSSQVLLITLINTERRHPGFATFFQLPRMSSCGRLKAKQGTNSPYPGHY 358

Qy 361 PPNIDCTWNIIEVPNNQHVVKFKFYLLLEPGVPACTCPKDYVEINGEKYCGERSQFVVT 420
Db 359 PPNIDCTWNIIEVPNNQHVVKFKFYLLLEPGVPACTCPKDYVEINGEKYCGERSQFVVT 418
Qy 421 NSNKITVRFHSDQSYTDTGFLAELYSYDSSDPCPQQTCTRTGRCIRKELRCDGWADCTDH 480
Db 419 NSNKITVRFHSDQSYTDTGFLAELYSYDSSDPCPQQTCTRTGRCIRKELRCDGWADCTDH 478
Qy 481 SDELNCSDAGHQTCKNFKCKPLFWVCDSDVNDGDNDSDEQCCSPAOTFRCSNGKCLSK 540
Db 479 SDELNCSDAGHQTCKNFKCKPLFWVCDSDVNDGDNDSDEQCCSPAOTFRCSNGKCLSK 538
Qy 541 SQCCNGKDDCGDSDEASCPKVNVTCTKHYRCLNGCLSKGNPECDKEDCSGDSBK 600
Db 539 SQCCNGKDDCGDSDEASCPKVNVTCTKHYRCLNGCLSKGNPECDKEDCSGDSBK 598
Qy 601 DDCGCLRSFTRQARVVGTTDADEGEWPMQVSLHALGQGHICGASLISPNMLVSAACHYD 660
Db 599 DDCGCLRSFTRQARVVGTTDADEGEWPMQVSLHALGQGHICGASLISPNMLVSAACHYD 658
Qy 661 DRGFRYSDPTQMTAFGLHLDQSQSAPGVOERRLKRIISHPPFNFTFDYDIALLELEKP 720
Db 659 DRGFRYSDPTQMTAFGLHLDQSQSAPGVOERRLKRIISHPPFNFTFDYDIALLELEKP 718
Qy 721 AYSVMVRPCLPDASHVFPAGKAIWVTGWGHTQVGGTGCALILQKEIRVINQTCENLL 780
Db 719 AYSVMVRPCLPDASHVFPAGKAIWVTGWGHTQVGGTGCALILQKEIRVINQTCENLL 778
Qy 781 PQQITPRMCMVGFSLGGVDSQCGDGGPLSSVEADGRIFQAGVSWGDCGCAQRNKPQVYT 840
Db 779 PQQITPRMCMVGFSLGGVDSQCGDGGPLSSVEADGRIFQAGVSWGDCGCAQRNKPQVYT 838
Qy 841 RLPLFRDWIKENTGV 855
Db 839 RLPLFRDWIKENTGV 853

RESULT 15
US-10-450-763-51801
; Sequence 51801, Application US/10450763
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 51801
; LENGTH: 932
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (715)..(733)
; OTHER INFORMATION: Kringle domain proteins domain identified by eMATRIX,
; OTHER INFORMATION: accession number BL00021B, p-value=5.696e-17, raw score of 13.3
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(932)
; OTHER INFORMATION: Xaa = X or * as defined in Table 2
US-10-450-763-51801

Query Match 92.3%; Score 4319; DB 6; Length 932;
Best Local Similarity 96.0%; Pred. No. 2,2e-313;
Matches 800; Conservative 7; Mismatches 22; Indels 4; Gaps 4;

```

QY 27 EKVNGLEGEVEFLPVNNVKKVEKHGGRWVLAALVIGLLLVLLGIGFLVWHLQYRDVRV 86
Db :|||||
QY 100 QKVNGLEGEVEFLPVNNVKKVEKHGGRWVLAALVIGLLLVLLGIGFLVWHLQYRDVRV 159
Db :|||||
QY 87 QKVFNQYMRITNENFVDAYENSNSFVSLSKVKDALKLLYSQVFFLGPYHKESAVTAF 146
Db :|||||
QY 160 QKVFNQYMRITNENFVDAYENSNSFVSLSKVKDALKLLYSQVFFLGPYHKESAVTAF 219
Db :|||||
QY 147 SEGSIAYYWSFSPQHLVEEAERVMAEERVMMLPPRARSLSKFVTSVVAFPDTSKTIV 206
Db :|||||
QY 220 SEGSIAYYWSFSPQHLVEEAERVMAEERVMMLPPRARSLSKFVTSVVAFPDTSKTIV 279
Db :|||||
QY 207 QRTQNSCSFGLHARGVELMRTTTFPGFDPSPYPAHARQWALRGDADSVLSITFRSFDLA 266
Db :|||||
QY 280 QRTQNSCSFGLHARGVELMRTTTFPGFDPSPYPAHARQWALRGDADSVLSITFRSFDLA 339
Db :|||||
QY 267 SCDEGRSDLVVTYNTLSMPHEHALVQLCGTYPPSYNLTFHSSQNVLLITLINTERRHPG 326
Db :|||||
QY 340 SCDEGRHLVVTYNTLSMPHEHALVQLCGTYPPSYNLTFHSSQNVLLITLINTERRHPG 399
Db :|||||
QY 327 PEATFQOLPRMSSCGRLRKAQGTNSPYYPGHYPP-NIDCTWNIENPNNOHVKVSFKPF 385
Db :|||||
QY 400 FEATFQOLPRMSSCGRLRKAQGTNSPYYPGHYPPQHXSTWNIEVENNOHVKVRFAFP 459
Db :|||||
QY 386 YLLEFGVPAGTCKDYVEINGEKYCGERSQFVVTSSNSNKITVRFHSDQSYTDTGFLAEYL 445
Db :|||||
QY 460 YLLEFGVPAGTCKDYVEINGEKYCGERSQFVVTSSNSNKITVRFHSDQSYTDTGFLAEYL 519
Db :|||||
QY 446 SYDSSDPCPGQFTCTGR-CIRKELRCQWADCTDHSDELNCSCDAGHQFTCKNFKCKPL 504
Db :|||||
QY 520 SYDSSDPCPGQFTCPHGAWVIRKELRCQWADCTDHSDELNCSCDAGHQFTCKNFKCKPL 579
Db :|||||
QY 505 FWVCDVNDGDNDSDEQSCPAQTFRCNSGKCLSKSQCNKDKDCGSDGSEASCPCVNV 564
Db :|||||
QY 580 FWVCDVNEGDNDSDEQSCPAQTFRCNSGKCLSKSQCNKDKDCGSDGSEASCPCVNV 639
Db :|||||
QY 565 VTCTKHTYRCNLGLCLSKGNPECDGKEDCGSDGSEKDCDGLRSFTRQARVVGTDADG 624
Db :|||||
QY 640 VTCTKHTYRCNLGLCLSKGNPECDGKEDCGSDGSEKDCDGLRSFTRQARVVGTDADG 699
Db :|||||
QY 625 EWPQVSLHALGQGHICGASLISPNWLVSAAHCYIDDRFRYSFDPQWT-AFLGLHDQSQ 683
Db :|||||
QY 700 EWPQVSLHALGQGHICGASLISPNWLVSAAHCFIDDRFRYSFDPQWT-AFLGLHDQSQ 759
Db :|||||
QY 684 RSAP-GVOERRLKRISHPFFNDFTDYDIALLELEKPAEYSSMVRPCLPDASHVFPAG 742
Db :|||||
QY 760 RPPWGCRRRLKRISHPFFNDFTDYDIALLELEKPAEYSSMVRPCLPDASHVFPAG 819
Db :|||||
QY 743 KAIWVTGHTQYGGTGALILQKGEIRVINQTTNENLLPQOITPRMCMCVGLSGGVDSQC 802
Db :|||||
QY 820 KAIWVTGHTQYGGTGALILQKGEIRVINQTTNENLLPQOITPRMCMCVGLSGGVDSQC 879
Db :|||||
QY 803 GDSGGPLSSVEADGRIFQAGVWSWGDGCAQRNKPQVYTRLPFLFRDWIKENTGV 855
Db :|||||
QY 880 GDSGGPLSSVEADGRIFQAGVWSWEGACVQRNKPQVYTRLPFLFRDWIKENTGV 932

```

Search completed: September 23, 2005, 13:09:19
Job time : 92 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 23, 2005, 12:50:11 ; Search time 26 Seconds
(without alignments)
2454.806 Million cell updates/sec

Title: US-09-421-213-2
Perfect score: 4681
Sequence: 1 MGSDDARKGGGPKDFGAGL.....PGVYTRPLFRDWIKENTGV 855

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents, AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4681	100.0	855	2	US-09-027-337-2
2	4681	100.0	855	4	US-09-644-600-2
3	4681	100.0	855	4	US-09-654-600A-2
4	3810	81.4	902	4	US-09-644-600-10
5	3810	81.4	902	4	US-09-654-600A-10
6	1319	28.2	241	4	US-09-657-986B-2
7	703.5	15.0	798	1	US-08-200-900A-2
8	703.5	15.0	798	4	US-08-794-042-2
9	703.5	15.0	798	5	PCT-US94-00616-2
10	692	14.8	1042	4	US-09-959-392-2
11	664.5	14.2	1113	4	US-09-959-392-4
12	588	12.6	407	3	US-09-734-675-4
13	577	12.3	717	4	US-09-949-016-11182
14	574	12.3	699	4	US-09-949-016-6138
15	560.5	12.0	492	4	US-09-685-166A-895
16	560.5	12.0	492	4	US-09-879-792-14
17	560.5	12.0	492	4	US-09-679-426-895
18	560.5	12.0	492	4	US-09-759-143-895
19	558.5	11.9	333	3	US-09-759-143-934
20	558.5	11.9	492	3	US-09-342-749-2
21	558.5	11.9	492	4	US-09-691-840-2
22	558.5	11.9	492	4	US-09-759-143-932
23	558.5	11.9	510	4	US-09-949-016-11074
24	549.5	11.7	232	4	US-09-959-392-32
25	547.5	11.7	235	3	US-08-944-483-65
26	544.5	11.6	235	3	US-08-807-151-3
27	544.5	11.6	235	3	US-09-478-957-3

28	543.5	11.6	454	3	US-09-518-046-2	Sequence 2, Appli
29	533	11.4	521	4	US-09-949-016-11081	Sequence 11081, A
30	533	11.4	521	4	US-09-949-016-11082	Sequence 11082, A
31	533	11.4	521	4	US-09-949-016-11083	Sequence 11083, A
32	527	11.3	446	4	US-10-177-661-4	Sequence 4, Appli
33	521.5	11.1	477	4	US-10-177-661-6	Sequence 2, Appli
34	518.5	11.1	477	4	US-10-177-661-2	Sequence 6, Appli
35	518.5	11.1	562	4	US-09-879-792-12	Sequence 12, Appli
36	514.5	11.0	638	2	US-08-681-151-3	Sequence 3, Appli
37	512.5	10.9	283	3	US-08-807-151-1	Sequence 1, Appli
38	512.5	10.9	283	3	US-08-478-957-1	Sequence 1, Appli
39	511.5	10.9	418	1	US-08-508-448C-25	Sequence 25, Appli
40	511.5	10.9	418	4	US-09-370-838-82	Sequence 82, Appli
41	511.5	10.9	418	4	US-09-370-838-83	Sequence 83, Appli
42	511.5	10.9	418	4	US-09-854-133-82	Sequence 82, Appli
43	511.5	10.9	418	4	US-09-854-133-83	Sequence 83, Appli
44	510.5	10.9	418	4	US-09-370-838-62	Sequence 62, Appli
45	510.5	10.9	418	4	US-09-854-133-62	Sequence 62, Appli

ALIGNMENTS

RESULT 1
US-09-027-337-2
; Sequence 2, Application US/09027337B
; Patent No. 5972616
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Tanimoto, Hirotooshi
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease Overexpressed in
; TITLE OF INVENTION: Breast and Ovarian Carcinomas
; FILE REFERENCE: D6064
; CURRENT APPLICATION NUMBER: US/09/027,337B
; CURRENT FILING DATE: 1998-02-20
; NUMBER OF SEQ ID NOS: 13
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Amino acid sequence of TADG-15 encoded by nucleotides
; OTHER INFORMATION: 23 to 2589 of Sequence 1
; Patent No. 5972616
US-09-027-337-2

Query Match	100.0%;	Score	4681;	DB	2;	Length	855;
Best Local Similarity	100.0%;	Pred.	No. 0;				
Matches	855;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						
Qy	1	MGSDDARKGGGPKDFGAGLKYNSRHEKVNGLSEGVFLPVNNVKYKKGPGRAWVLLAA	60				
Db	1	MGSDDARKGGGPKDFGAGLKYNSRHEKVNGLSEGVFLPVNNVKYKKGPGRAWVLLAA	60				
Qy	61	VLIGLLLVLLIGLFLVHLLQYRDVRVQKVFNGYMRITNENFVDAYENSNSFEVSLASKV	120				
Db	61	VLIGLLLVLLIGLFLVHLLQYRDVRVQKVFNGYMRITNENFVDAYENSNSFEVSLASKV	120				
Qy	121	KDALKLLYSVPFLGPPYHKESAVTASEGSIYAYNSSEPSIPQHLVEAEERVAEERVVM	180				
Db	121	KDALKLLYSVPFLGPPYHKESAVTASEGSIYAYNSSEPSIPQHLVEAEERVAEERVVM	180				
Qy	181	LPFRARSLKSFVTSVVAFFPTDSKTORTODNCSFGLHARGVBLMFTTTPGFPDPSYPYA	240				
Db	181	LPFRARSLKSFVTSVVAFFPTDSKTORTODNCSFGLHARGVBLMFTTTPGFPDPSYPYA	240				
Qy	241	HARQWALRGDADSVLSLTFERSFDLASCDERGSDLVTVTNTLSPMEHALVQLCGTTPPS	300				
Db	241	HARQWALRGDADSVLSLTFERSFDLASCDERGSDLVTVTNTLSPMEHALVQLCGTTPPS	300				
Qy	301	YNLTFHSSQNVLITITNTERRHPGFEATFFQLPFRMSSCGGLRKAQGTNSPYYPGHY	360				
Db	301	YNLTFHSSQNVLITITNTERRHPGFEATFFQLPFRMSSCGGLRKAQGTNSPYYPGHY	360				

361 PPNICTWNIIEVPNNQHVSKFVYLLLEPGVPAGTCKPKDYVEINGEKYCGERSQFVVT 420
 361 PPNICTWNIIEVPNNQHVSKFVYLLLEPGVPAGTCKPKDYVEINGEKYCGERSQFVVT 420
 421 NSNKITVRFSHDSQSYTDTGFLAELYSYDSDDPCPGQFTCTGRCIRKELRCGWDADCTDH 480
 421 NSNKITVRFSHDSQSYTDTGFLAELYSYDSDDPCPGQFTCTGRCIRKELRCGWDADCTDH 480
 481 SDELNCSADAGHFTCKNFKCKPLFWVCDSDVNDGNSDEQSCPAQTFRCNSGKCLSK 540
 481 SDELNCSADAGHFTCKNFKCKPLFWVCDSDVNDGNSDEQSCPAQTFRCNSGKCLSK 540
 541 SQQCKGKDCGDSDEASCPKVVVCTCTKTYRCLNGLCLSKGNPECDGKEDCSGSD 600
 541 SQQCKGKDCGDSDEASCPKVVVCTCTKTYRCLNGLCLSKGNPECDGKEDCSGSD 600
 601 DDCGLRSFTRQARVVGTDADGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
 601 DDCGLRSFTRQARVVGTDADGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
 661 DRGFYSDPTQWTAFLGLHDQSORSAFGVQERRLKRIISHPPFNDFTPDYDIALLELEKP 720
 661 DRGFYSDPTQWTAFLGLHDQSORSAFGVQERRLKRIISHPPFNDFTPDYDIALLELEKP 720
 721 AYEYSSMRVPCICLPDASHVFPAGKAIWVTGWGHTQYGGTGALLILOKGEIRVINQTTCE 780
 721 AYEYSSMRVPCICLPDASHVFPAGKAIWVTGWGHTQYGGTGALLILOKGEIRVINQTTCE 780
 781 PQQITPRMNCVGLSGVDSCQDGGPLSSVEADGRIFQAGVSWGDCQAQRNKP 840
 781 PQQITPRMNCVGLSGVDSCQDGGPLSSVEADGRIFQAGVSWGDCQAQRNKP 840
 841 RLPLFRDWIKENTGV 855
 841 RLPLFRDWIKENTGV 855

RESULT 2
 US-09-644-600-2
 ; Sequence 2, Application US/09644600
 ; Patent No. 6451500
 ; GENERAL INFORMATION:
 ; APPLICANT: O'Brien, Timothy J.
 ; APPLICANT: Tanimoto, Hirotochi
 ; TITLE OF INVENTION: TAGD-15: An Extracellular Serine Protease
 ; FILE OF INVENTION: Overexpressed in Carcinomas
 ; FILE REFERENCE: D6064CIP/D
 ; CURRENT APPLICATION NUMBER: US/09/644,600
 ; CURRENT FILING DATE: 2000-08-23
 ; PRIOR APPLICATION NUMBER: 09/421,213
 ; PRIOR FILING DATE: 1999-10-20
 ; PRIOR APPLICATION NUMBER: 09/027,337
 ; PRIOR FILING DATE: 1998-02-20
 ; NUMBER OF SEQ ID NOS: 98
 ; SEQ ID NO 2
 ; LENGTH: 855
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: TAGD-15
 US-09-644-600-2

Query Match 100.0%; Score 4681; DB 4; Length 855;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 855; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 1 MGSDDRKGGGKDFGAGLKYNSRHEKYNGLGEEVFLPVNNVKKVEKHGGRVVLAA 60
 1 MGSDDRKGGGKDFGAGLKYNSRHEKYNGLGEEVFLPVNNVKKVEKHGGRVVLAA 60
 61 VLIGLLLVLLGIGFLVHLOYRDVRVQKVFNGYMRITNENFVDAYENSNSTEFVSLASKV 120

Db 61 VLIGLLLVLLGIGFLVHLOYRDVRVQKVFNGYMRITNENFVDAYENSNSTEFVSLASKV 120
 QY 121 KDALKLLYSGVPFFLGPYHKSATVAFSGSVIAYYWSFSPQHLVEBAERVMASERVVM 180
 Db 121 KDALKLLYSGVPFFLGPYHKSATVAFSGSVIAYYWSFSPQHLVEBAERVMASERVVM 180
 QY 181 LPPRARSLSKSPVVTSSVAPPTDSKITVQRTQDNSCSFGLHARGVELMRFTTGGFPDPSYP 240
 Db 181 LPPRARSLSKSPVVTSSVAPPTDSKITVQRTQDNSCSFGLHARGVELMRFTTGGFPDPSYP 240
 QY 241 HARQWALRGDADSVLSITFRSPDLASCDERGSDDLVTYNTLSMPHEHALVOLCCTYPPS 300
 Db 241 HARQWALRGDADSVLSITFRSPDLASCDERGSDDLVTYNTLSMPHEHALVOLCCTYPPS 300
 QY 301 YNLTFHSSQNVLLITLINTERRHPGFEATFQQLPRMSSCGRLRKAQGTNSPYYPGHY 360
 Db 301 YNLTFHSSQNVLLITLINTERRHPGFEATFQQLPRMSSCGRLRKAQGTNSPYYPGHY 360
 QY 361 PPNICTWNIIEVPNNQHVSKFVYLLLEPGVPAGTCKPKDYVEINGEKYCGERSQFVVT 420
 Db 361 PPNICTWNIIEVPNNQHVSKFVYLLLEPGVPAGTCKPKDYVEINGEKYCGERSQFVVT 420
 QY 421 NSNKITVRFSHDSQSYTDTGFLAELYSYDSDDPCPGQFTCTGRCIRKELRCGWDADCTDH 480
 Db 421 NSNKITVRFSHDSQSYTDTGFLAELYSYDSDDPCPGQFTCTGRCIRKELRCGWDADCTDH 480
 QY 481 SDELNCSADAGHFTCKNFKCKPLFWVCDSDVNDGNSDEQSCPAQTFRCNSGKCLSK 540
 Db 481 SDELNCSADAGHFTCKNFKCKPLFWVCDSDVNDGNSDEQSCPAQTFRCNSGKCLSK 540
 QY 541 SQQCKGKDCGDSDEASCPKVVVCTCTKTYRCLNGLCLSKGNPECDGKEDCSGSD 600
 Db 541 SQQCKGKDCGDSDEASCPKVVVCTCTKTYRCLNGLCLSKGNPECDGKEDCSGSD 600
 QY 601 DDCGLRSFTRQARVVGTDADGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
 Db 601 DDCGLRSFTRQARVVGTDADGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
 QY 661 DRGFYSDPTQWTAFLGLHDQSORSAFGVQERRLKRIISHPPFNDFTPDYDIALLELEKP 720
 Db 661 DRGFYSDPTQWTAFLGLHDQSORSAFGVQERRLKRIISHPPFNDFTPDYDIALLELEKP 720
 QY 721 AYEYSSMRVPCICLPDASHVFPAGKAIWVTGWGHTQYGGTGALLILOKGEIRVINQTTCE 780
 Db 721 AYEYSSMRVPCICLPDASHVFPAGKAIWVTGWGHTQYGGTGALLILOKGEIRVINQTTCE 780
 QY 781 PQQITPRMNCVGLSGVDSCQDGGPLSSVEADGRIFQAGVSWGDCQAQRNKP 840
 Db 781 PQQITPRMNCVGLSGVDSCQDGGPLSSVEADGRIFQAGVSWGDCQAQRNKP 840
 QY 841 RLPLFRDWIKENTGV 855
 Db 841 RLPLFRDWIKENTGV 855

RESULT 3
 US-09-654-600A-2
 ; Sequence 2, Application US/09654600A
 ; Patent No. 6649741
 ; GENERAL INFORMATION:
 ; APPLICANT: O'Brien, Timothy J.
 ; APPLICANT: Tanimoto, Hirotochi
 ; TITLE OF INVENTION: TAGD-15: An Extracellular Serine Protease
 ; FILE OF INVENTION: Overexpressed in Carcinomas
 ; FILE REFERENCE: D6064CIP/D
 ; CURRENT APPLICATION NUMBER: US/09/654,600A
 ; CURRENT FILING DATE: 2000-09-01
 ; PRIOR APPLICATION NUMBER: 09/421,213
 ; PRIOR FILING DATE: 1999-10-20
 ; NUMBER OF SEQ ID NOS: 98
 ; SEQ ID NO 2

```
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: TADG-15
US-09-654-600A-2

Query Match      100.0%; Score 4691; DB 4; Length 855;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 855; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGS DRARKGGGKPGAGLKYNSRHEKNGLEEGVEFLPVNNKKVEKHGPRWVLLAA 60
Db 1 MGS DRARKGGGKPGAGLKYNSRHEKNGLEEGVEFLPVNNKKVEKHGPRWVLLAA 60
QY 61 VLIGLLLVLLGIGFLVWHLYQYRDVVRQKVFNGYMRITNENFVDAYENSNSSTFVSLASKV 120
Db 61 VLIGLLLVLLGIGFLVWHLYQYRDVVRQKVFNGYMRITNENFVDAYENSNSSTFVSLASKV 120
QY 121 KDALKLLYSGVPLFLPGYHKESAVTAFSEGSVIAYYWSFSPHQHLYVEAEARVMAERVVM 180
Db 121 KDALKLLYSGVPLFLPGYHKESAVTAFSEGSVIAYYWSFSPHQHLYVEAEARVMAERVVM 180
QY 181 LPPRARSLSKSFVTSVVAEPTDSKTORTODNSCSFGLHARGVELMRFTTPGPPSPYPA 240
Db 181 LPPRARSLSKSFVTSVVAEPTDSKTORTODNSCSFGLHARGVELMRFTTPGPPSPYPA 240
QY 241 HARCOWALRGDADSVLSLTFRSFDLASCDERGSDDLVTYNTLSPMEPHALVOLCGTYPPS 300
Db 241 HARCOWALRGDADSVLSLTFRSFDLASCDERGSDDLVTYNTLSPMEPHALVOLCGTYPPS 300
QY 301 YNLTFHSSQNVLLITLITNTERRHPGFEATFQLPKMSSCGGLRKAQGTFFNSPYYPGHY 360
Db 301 YNLTFHSSQNVLLITLITNTERRHPGFEATFQLPKMSSCGGLRKAQGTFFNSPYYPGHY 360
QY 361 PNIDCTWNIIEVPNNQHVKSFKFYLLEPGVPAGTCKPDYVEINGEKYCGERSQFVTS 420
Db 361 PNIDCTWNIIEVPNNQHVKSFKFYLLEPGVPAGTCKPDYVEINGEKYCGERSQFVTS 420
QY 421 NSNKITVRPHSDQSYTDTGFLAEYLSYSDSDPCPGQFTCTGRCIRKELRCGWDADCTDH 480
Db 421 NSNKITVRPHSDQSYTDTGFLAEYLSYSDSDPCPGQFTCTGRCIRKELRCGWDADCTDH 480
QY 481 SDELNCSCDAGHQTCKNKFCKPLFWVCDSDVNDGDNDSDEQGCSPAQTFRCNSGKCLSK 540
Db 481 SDELNCSCDAGHQTCKNKFCKPLFWVCDSDVNDGDNDSDEQGCSPAQTFRCNSGKCLSK 540
QY 541 SQCCNGKDDCGDSDGDEASCPKVNVTCTKHTYRCLNGLCLSKGNPECDGKDCSDGDEK 600
Db 541 SQCCNGKDDCGDSDGDEASCPKVNVTCTKHTYRCLNGLCLSKGNPECDGKDCSDGDEK 600
QY 601 DCDCLGSRSTRQARVVGTTDADGEGWPMQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
Db 601 DCDCLGSRSTRQARVVGTTDADGEGWPMQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
QY 661 DRGFRYSDPTQWTAFLGLHDQSORAPGVQERLKRILISHPFNFDFDYDIALLELEKP 720
Db 661 DRGFRYSDPTQWTAFLGLHDQSORAPGVQERLKRILISHPFNFDFDYDIALLELEKP 720
QY 721 ABYSNMVRPICLPDASHVPPAGKAIWVTGWGHTQYGGTGALILQKEIRVINQTTCEMLL 780
Db 721 ABYSNMVRPICLPDASHVPPAGKAIWVTGWGHTQYGGTGALILQKEIRVINQTTCEMLL 780
QY 781 PQOITPRMNCVGLSGGVDS CGDGGPLSSVEADGRIFQAGVSWGDCGCAQRNKGYYT 840
Db 781 PQOITPRMNCVGLSGGVDS CGDGGPLSSVEADGRIFQAGVSWGDCGCAQRNKGYYT 840
QY 841 RLPLFRDWIKENTGV 855
Db 841 RLPLFRDWIKENTGV 855
```

```
US-09-644-600-10
; Sequence 10, Application US/09644600
; Patent No. 6451500
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Tanimoto, Hirotochi
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease
; FILE REFERENCE: D6064CIP/D
; CURRENT APPLICATION NUMBER: US/09/644,600
; CURRENT FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: 09/421,213
; PRIOR FILING DATE: 1999-10-20
; PRIOR APPLICATION NUMBER: 09/027,337
; PRIOR FILING DATE: 1998-02-20
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 10
; LENGTH: 902
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
; OTHER INFORMATION: Epithin
US-09-644-600-10

Query Match      81.4%; Score 3810; DB 4; Length 902;
Best Local Similarity 81.4%; Pred. No. 1.1e-270;
Matches 686; Conservative 71; Mismatches 86; Indels 0; Gaps 0;

QY 1 MGS DRARKGGGKPGAGLKYNSRHEKNGLEEGVEFLPVNNKKVEKHGPRWVLLAA 60
Db 1 MGS NGRKAGGSGQDFGAGLYKDYSLNMGVEEGVEFLPANNKKVEKGRPRWVLLVA 60
QY 61 VLIGLLLVLLGIGFLVWHLYQYRDVVRQKVFNGYMRITNENFVDAYENSNSSTFVSLASKV 120
Db 61 VLFSFLLLSLWAGLLVWHFYRNVRQKVFNGHLRITNEIFLDAYENSTSTFISLASQV 120
QY 121 KDALKLLYSGVPLFLPGYHKESAVTAFSEGSVIAYYWSFSPHQHLYVEAEARVMAERVVM 180
Db 121 KEALKLLYNEVPLGVPHYKKSAVTAFSEGSVIAYYWSFSPHQHLYVEAEARVMAERVVM 180
QY 181 LPPRARSLSKSFVTSVVAEPTDSKTORTODNSCSFGLHARGVELMRFTTPGPPSPYPA 240
Db 181 LPPRARSLSKSFVTSVVAEPTDSKTORTODNSCSFGLHARGVELMRFTTPGPPSPYPA 240
QY 241 HARCOWALRGDADSVLSLTFRSFDLASCDERGSDDLVTYNTLSPMEPHALVOLCGTYPPS 300
Db 241 HARCOWALRGDADSVLSLTFRSFDVAPCDEHGSDDLVTYDLSLSPMEHAYVVLCTGTFSS 300
QY 301 YNLTFHSSQNVLLITLITNTERRHPGFEATFQLPKMSSCGGLRKAQGTFFNSPYYPGHY 360
Db 301 YNLTFHSSQNVLLITLITNTERRHPGFEATFQLPKMSSCGGLRKAQGTFFNSPYYPGHY 360
QY 361 PNIDCTWNIIEVPNNQHVKSFKFYLLEPGVPAGTCKPDYVEINGEKYCGERSQFVTS 420
Db 361 PNINCTWNIIPNNRNKVKRFLFYLDVDPNVPVGSCTKYVEINGEKSGERSQFVSS 420
QY 421 NSNKITVRPHSDQSYTDTGFLAEYLSYSDSDPCPGQFTCTGRCIRKELRCGWDADCTDH 480
Db 421 NSSKLTIVHSDHSYTDGFLAEYLSYSDSDPCPGFMCKTGRCIRKELRCGWDADCTDH 480
QY 481 SDELNCSCDAGHQTCKNKFCKPLFWVCDSDVNDGDNDSDEQGCSPAQTFRCNSGKCLSK 540
Db 481 SDERYCRNATHQFTCKNQFCKPLFWVCDSDVNDGDSDEGSCSPAGSKCSNGKCLPQ 540
QY 541 SQCCNGKDDCGDSDGDEASCPKVNVTCTKHTYRCLNGLCLSKGNPECDGKDCSDGDEK 600
Db 541 SQKCNKDNCGDSDGDEASCDVNVVSCCTKYTRCQNGCLCLSKGNPECDGKDCSDGDEK 600
QY 601 DCDCLGSRSTRQARVVGTTDADGEGWPMQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
Db 601 NCDCLGSRSTRQARVVGTTDADGEGWPMQVSLHALGQGHICGASLISFDWLVSAAHCQD 660
QY 661 DRGFRYSDPTQWTAFLGLHDQSORAPGVQERLKRILISHPFNFDFDYDIALLELEKP 720
```

Db 661 DKNFKYSDYTWAFGLDQSRSSAGVQELKRIITHPSFNDFDYDIALLELEKS 720
QY 721 AYSVMVRPCLPDASHVFPAGKAIWVTGHTQYGTGALILQKGEIRVINQTTCEMLL 780
Db 721 VEYSTVVRPCLPDATHVFPAGKAIWVTGHTKEGTTGALILQKGEIRVINQTTCEMLL 780
QY 781 PQOITPRMCMVGLSGGVDSQCGDGLSSVSEADGRIFQAGVSVSGDGCQAQRNKGVT 840
Db 781 PQOITPRMCMVGLSGGVDSQCGDGLSSVSEADGRIFQAGVSVSGDGCQAQRNKGVT 840
QY 841 RLP 843
Db 841 RLP 843

RESULT 5

US-09-654-600A-10
; Sequence 10, Application US/09654600A
; Patent No. 6649741
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Tanimoto, Hirotsoshi
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease
; FILE REFERENCE: Overexpressed in Carcinomas
; CURRENT APPLICATION NUMBER: US/09/654,600A
; CURRENT FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 09/421,213
; 09/027,337
; PRIOR FILING DATE: 1999-10-20
; 1998-02-20
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 10
; LENGTH: 902
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
; OTHER INFORMATION: Epithin
US-09-654-600A-10

Query Match 81.4%; Score 3810; DB 4; Length 902;
Best Local Similarity 81.4%; Pred. No. 1.1e-270;
Matches 686; Conservative 71; Mismatches 86; Indels 0; Gaps 0;

QY 1 MGSDRARKGGGKDFGAGLKYNRHEKVNLEEGVEFLPVANNVKKVKGPRVWLAA 60
Db 1 MGSNRGRKAGGGQDFGAGLKYSRLNNGVEEGVEFLPVANNVKKVKGPRVWLVA 60
QY 61 VLIGLLVLLGIFLYWHLQYRDVRVQKVFNGYMRITNENFVDAYENSSTFEVSLASKV 120
Db 61 VLFSFLLLSLMAELLVWHPHYRNVQKVFNGHLRITNEIFLDAYENSTSTEFISLQV 120
QY 121 KQALKLLYSGVPLPGYHESAVTAPSEGSVAYYWSFISIPQHLVEAEERVAERVVM 180
Db 121 KEALKLLYNEVPVLPGYHKSAVTAPEGSVAYYWSFISIPPLAEYEDRAMAVERVYT 180
QY 181 LPPARSLKSFVYTSVVAFPTDSKTQRTQDNCSPGLHARGVLMRFTTPGFPDPSYPA 240
Db 181 LPPARALKSFVYTSVVAFIDPRLMORTQDNCSPGLHARGVLMRFTTPGFPDPSYPA 240
QY 241 HARCQWALRGDADSVLSLTFRSDLASCDERGSDDLVTYNTLSMPPEHALVOLCGTTPPS 300
Db 241 HARCQWALRGDADSVLSLTFRSDVAPCDHEGSDLVTVYDLSLMPPEHALVOLCGTTPPS 300
QY 301 YNLTFHSSQNVLLITLITTRRHQGFATFPQLPRMSSCGGLRKAQOTFNSPYYPGHY 360
Db 301 YNLTFHSSQNVLLITLITNTRHGLGFATFPQLPRMSSCGGLRKAQOTFNSPYYPGHY 360
QY 361 PNINDCTWNIENVPNNOHVKSFKFFVLLPEGPVAGTCPKDYVEINKEKYCGRSQFVTS 420
Db 361 PNINDCTWNIENVPNNRNVKFKLFYLVDPNVVSGCTKDYVEINKEKSGRSQFVSS 420

QY 421 NSNKITVRFHSDQSYTDGFLAEYLSYSDSPCPGQQTCTRGRCIRKELRCGWDGADCTDH 480
Db 421 NSSKITVRFHSDHSYTDGFLAEYLSYSDNDPCPGMFMCKTGRCIRKELRCGWDGADCPDY 480
QY 481 SDELNCSADAGHOFTCKKFKCKPLFWVCDSDVNDGDNDEQCCSCPAQOTFRCSNGKCLSK 540
Db 481 SDERYCRNATHQFTCKNQFCCKPLFWVCDSDVNDGDSDEGSCSPAGSKCSNGKCLPQ 540
QY 541 SOQCNKDDCGDSDGDEASCPKVVVVTCTKHYRCLNGLCLSKNGNPECDGKEDCSGSDRK 600
Db 541 SQKCNKDNCGDSDGDEASCDVNVVSTKYRRCQNGLCCLSKNGNPECDGKEDCSGSDRK 600
QY 601 DDCGLRSFTRQARVVGCTDADEGEWPMQVSLHALGQGHICGASLISPNMLVSAACHYID 660
Db 601 NDCGLRSFTRQARVVGCTNADEGEWPMQVSLHALGQGHICGASLISPDWLVSAACHFOD 660
QY 661 DRGFRYSDPTOWTAFGLHDQSORAPGVQRRRLKRIISHPFNFDFDYDIALLELEKP 720
Db 661 DKNFKYSDYTWAFGLDQSRSSAGVQELKRIITHPSFNDFDYDIALLELEKS 720
QY 721 AYSVMVRPCLPDASHVFPAGKAIWVTGHTQYGTGALILQKGEIRVINQTTCEMLL 780
Db 721 VEYSTVVRPCLPDATHVFPAGKAIWVTGHTKEGTTGALILQKGEIRVINQTTCEMLL 780
QY 781 PQOITPRMCMVGLSGGVDSQCGDGLSSVSEADGRIFQAGVSVSGDGCQAQRNKGVT 840
Db 781 PQOITPRMCMVGLSGGVDSQCGDGLSSVSEADGRIFQAGVSVSGDGCQAQRNKGVT 840
QY 841 RLP 843
Db 841 RLP 843

RESULT 6

US-09-657-986B-2
; Sequence 2, Application US/09657986B
; Patent No. 6797504
; GENERAL INFORMATION:
; APPLICANT: Madison, Edwin L.
; APPLICANT: Semple, Joseph Edward
; APPLICANT: Coombs, Gary Samuel
; APPLICANT: Reiner, John Eugene
; APPLICANT: Ong, Edgar O.
; APPLICANT: Araldi, Gian Luca
; TITLE OF INVENTION: Inhibitors of Serine Protease Activity of Matriptase or
; TITLE OF INVENTION: MTSPI
; FILE REFERENCE: Corvas 255/049
; CURRENT APPLICATION NUMBER: US/09/657,986B
; CURRENT FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 241
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-09-657-986B-2

Query Match 28.2%; Score 1319; DB 4; Length 241;
Best Local Similarity 100.0%; Pred. No. 6.7e-89;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 615 VVGCTDADEGEWPMQVSLHALGQGHICGASLISPNMLVSAACHYIDDRGFRYSDPTOWTA 674
Db 1 VVGCTDADEGEWPMQVSLHALGQGHICGASLISPNMLVSAACHYIDDRGFRYSDPTOWTA 60
QY 675 FLGLHDQSORAPGVQRRRLKRIISHPFNFDFDYDIALLELEKPAYSVSWVRPCLCPD 734
Db 61 FLGLHDQSORAPGVQRRRLKRIISHPFNFDFDYDIALLELEKPAYSVSWVRPCLCPD 120
QY 735 ASHVFPAGKAIWVTGHTQYGTGALILQKGEIRVINQTTCEMLLPOQITPRMCMVGL 794
Db 121 ASHVFPAGKAIWVTGHTQYGTGALILQKGEIRVINQTTCEMLLPOQITPRMCMVGL 180

QY 795 SGGVDSQCGSGPLSSVADGRIFQAGVVSMDGCAQRNKGQVYTRLPFLFRDWIKENTG 854
Db 181 SGGVDSQCGSGPLSSVADGRIFQAGVVSMDGCAQRNKGQVYTRLPFLFRDWIKENTG 240
QY 855 V 855
Db 241 V 241

RESULT 7

US-08-200-900A-2
; Sequence 2, Application US/08200900A
; Patent No. 5665566
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc. - Legal Affairs
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/200,900A
; FILING DATE: 23-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Meinhert, Maureen C.
; REGISTRATION NUMBER: 31,544
; REFERENCE/DOCKET NUMBER: GI 5201-FWC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 876-1170 X8574
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 798 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-200-900A-2

Query Match 15.0%; Score 703.5; DB 1; Length 798;
Best Local Similarity 32.6%; Pred. No. 3.8e-43;
Matches 181; Conservative 91; Mismatches 187; Indels 97; Gaps 24;
QY 333 QLPRMSSCGG--RLRKAQGTFSPPYGHYPNIDCTNIEVPNNQHVKSFKFFYLLEP 390
Db 298 ELP--TDCGGPHDLWEPTTFTSINFPNSYPNQAFCIWNLNAQKGNQLHFEFDLENI 355
QY 391 GVPAGTCKDYVEI--NGEKYCGERSQFV-----VTSNKNKITVRFHSDQSYTDT 438
Db 356 A-----DVVEIRDE---GDSLFLAVYTGPGPVNDVFTNRTMTVLTIDNMLAQ 404
QY 439 GLAEVLS---YDSDPC--PGQFTCTGTCIRKELRCDCGWADCTHSDLENC-----SC 488
Db 405 GFKANFTTGYGLGIPCEKEDNFQCKDGCIPLVNLCDFGPHCKDGSDEAHCVRFLNGTT 464
QY 489 DAGH--QFTCKNFKCKPLFWVCDSDVNDGDSQSCPAOTFRCSNGKILSKSQOQNG 546
Db 465 DSSGLVQFIQS-----LWHV-----ACAEN-----WTTQISDDVC-----QLLG 499
QY 547 KDDCGDGSDEAC-----PKVNVVTCRTHYRCLNGLCLSKGNPECDGKEDCSGDEK 600
Db 500 ---LGTGNSVFPFGTGGGPPYVNLNTP-----NGSLILTPSQQC-----LEDSLILL 544

QY 601 DCD-----CGLRSFTQA--RVVGGTDADEGEWMPQVSLHALGQGHICGASLISPNMLVSA 655
Db 545 QCNVKSCKGLVTVQVSPKIVGGSDSREGAWPWWVALYFDDQ--QVCGASLVSRLMLVSA 603
QY 656 HCYIDDRGRFYSDPTQWTAFLGLHDQSQRSAFCVQERRLKRILSHPPFNDFTFDYDIAL 715
Db 604 HCVYG---RNMPSKWKAVLGLHNASNLTSQIETRLIDQIVINPHYNKRRKNNDIAMM 659
QY 716 ELEKPAEYSSMVRPICLPDASHVFPAGKAIWVTGHTQYGGTGALILQKGRVIRVNTT 775
Db 660 HLEKVVNTDYIOPICLPEENQVFPFGRICSIAGHGLIYQGSTADVLQEAADVPLLSNEK 719
QY 776 CENLPPQ-QITPRMCMVGLSGVDSQCGDGGPLSSVADGRIFQAGVVSMDGCAQRN 834
Db 720 CQQQMPYNTITENMVCAGYEAGGVDSQCGDGGPLMCQE--NNRWLLAGVTSFGYOCALPN 778
QY 835 KPGVYTRLPFLFRDWIK 850
Db 779 RPGVYARVPFTEWIK 794

RESULT 8

US-08-794-042-2
; Sequence 2, Application US/08794042
; Patent No. 6746859
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc. - Legal Affairs
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/794,042
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/200,900
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Meinhert, Maureen C.
; REGISTRATION NUMBER: 31,544
; REFERENCE/DOCKET NUMBER: GI 5201-FWC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 876-1170 X8574
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 798 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-794-042-2

Query Match 15.0%; Score 703.5; DB 4; Length 798;
Best Local Similarity 32.6%; Pred. No. 3.8e-43;
Matches 181; Conservative 91; Mismatches 187; Indels 97; Gaps 24;
QY 333 QLPRMSSCGG--RLRKAQGTFSPPYGHYPNIDCTNIEVPNNQHVKSFKFFYLLEP 390
Db 298 ELP--TDCGGPHDLWEPTTFTSINFPNSYPNQAFCIWNLNAQKGNQLHFEFDLENI 355
QY 391 GVPAGTCKDYVEI--NGEKYCGERSQFV-----VTSNKNKITVRFHSDQSYTDT 438

Db 356 A-----DVAIRIDGE---GDSLFLAVYTGPGVNDVFSTNRMVTLFIDNMLAQ 404
QY 439 GFLAEVLS---YDSSDP-PCQGTCTGRCIRKELRCDCGWADCTDHSDELNC-----SC 488
Db 405 GFKANFTTGYGLGIPBCKEDNFQCKDGEICPLVNLCDGPFPHCKDGSDEAHCVRLPFGTT 464
QY 489 DAGH--OFTCKNFKCKPLFWVCDSDVNDGDNDEQSCCPAQTFRCNSGKCLSKSQCCNG 546
Db 465 DSSGLVQFRIQS-----IWHV-----ACAEN-----WTTQISDDVC-----QLLG 499
QY 547 KDDCGDGSDEASC-----PKNVVTCYKHTYCLNGLCLSKGNPECDGKEDCSGSDK 600
Db 500 ---LGTGNSVPTFTGGGPPVNLNTAP-----NGSLILTPSOQC-----LEDSLILL 544
QY 601 DCD---CGLRSFTROA--RVVGTDDADGEGWPMOVSJLHALGQGHICGASLISPNMLVSA 655
Db 545 QCNYKSCGKCLVQEVSPKIVGSDSREGAPWVVALYFDDQ-QVCGASLVRDMLVSA 603
QY 656 HCVIDRGRFYSPTQWTAFLGLHDQSORSAPGVOERRLKRILSHPPFNDFTFDYDIAL 715
Db 604 HCIVG---RNMPEPSKWKAVLGLHMASNLTSPOIETRLIDQIVINPHYNKRKNNDIAM 659
QY 716 ELEKPAEYSSMRPCLPDASHVFPAGKAIWVTGHTQYGGTGALILQKGEIRVINQTT 775
Db 660 HLEKMYNTDYIQICLPEENQVFPFGRICSIAGWGLIYQGSTADVLQADVPPLSNEK 719
QY 776 CENLLPQ-QITPRMCMVGLSGVDSCQDGGPLSSVEADGRIFOAGVWSMGDCAQRN 834
Db 720 CQOQMEYNITENMVCAGYEAGGVDSQDGGPLMCQE--NNRWLLAGVTSFGYQCALPN 778
QY 835 KPGVYTRLPFRDWIK 850
Db 779 RGVYARVPRFTWIIQ 794

RESULT 9

PCT-US94-00616-2

; Sequence 2, Application PC/TUS9400616
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE
; NUMBER OF SEQUENCES: 33
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 798 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US94-00616-2

Query Match 15.0%; Score 703.5; DB 5; Length 798;
Best Local Similarity 32.6%; Pred. No. 3.8e-43;
Matches 181; Conservative 91; Mismatches 187; Indels 97; Gaps 24;
QY 333 QLPKMSGCG--RLRKAQTFTNSPYPGHYPPNIDCTNIEVPNNQHVKSFKPYLLEP 390
Db 298 ELF--TDCGGPHLEPNTFTTSINFPNSYPNQAFCIMNLNAQKQLQLHFQEPDENI 355
QY 391 GVPAGTCPKDYVEI--NGEKYCCERSQFV-----VTSNKNITVRHFSHQSYTDT 438
Db 356 A-----DVAIRIDGE---GDSLFLAVYTGPGVNDVFSTNRMVTLFIDNMLAQ 404

QY 439 GFLAEVLS---YDSSDP-PCQGTCTGRCIRKELRCDCGWADCTDHSDELNC-----SC 488
Db 405 GFKANFTTGYGLGIPBCKEDNFQCKDGEICPLVNLCDGPFPHCKDGSDEAHCVRLPFGTT 464
QY 489 DAGH--OFTCKNFKCKPLFWVCDSDVNDGDNDEQSCCPAQTFRCNSGKCLSKSQCCNG 546
Db 465 DSSGLVQFRIQS-----IWHV-----ACAEN-----WTTQISDDVC-----QLLG 499
QY 547 KDDCGDGSDEASC-----PKNVVTCYKHTYCLNGLCLSKGNPECDGKEDCSGSDK 600
Db 500 ---LGTGNSVPTFTGGGPPVNLNTAP-----NGSLILTPSOQC-----LEDSLILL 544
QY 601 DCD---CGLRSFTROA--RVVGTDDADGEGWPMOVSJLHALGQGHICGASLISPNMLVSA 655
Db 545 QCNYKSCGKCLVQEVSPKIVGSDSREGAPWVVALYFDDQ-QVCGASLVRDMLVSA 603
QY 656 HCVIDRGRFYSPTQWTAFLGLHDQSORSAPGVOERRLKRILSHPPFNDFTFDYDIAL 715
Db 604 HCIVG---RNMPEPSKWKAVLGLHMASNLTSPOIETRLIDQIVINPHYNKRKNNDIAM 659
QY 716 ELEKPAEYSSMRPCLPDASHVFPAGKAIWVTGHTQYGGTGALILQKGEIRVINQTT 775
Db 660 HLEKMYNTDYIQICLPEENQVFPFGRICSIAGWGLIYQGSTADVLQADVPPLSNEK 719
QY 776 CENLLPQ-QITPRMCMVGLSGVDSCQDGGPLSSVEADGRIFOAGVWSMGDCAQRN 834
Db 720 CQOQMEYNITENMVCAGYEAGGVDSQDGGPLMCQE--NNRWLLAGVTSFGYQCALPN 778
QY 835 KPGVYTRLPFRDWIK 850
Db 779 RGVYARVPRFTWIIQ 794

RESULT 10

US-09-959-392-2

; Sequence 2, Application US/09959392
; Patent No. 6806075
; GENERAL INFORMATION:
; APPLICANT: MORSE, MICHAEL JOHN
; APPLICANT: WU, QINGYU
; APPLICANT: YAN, WEI
; TITLE OF INVENTION: CORIN, A SERINE PROTEASE
; FILE REFERENCE: BERLX 74A
; CURRENT APPLICATION NUMBER: US/09/959,392
; CURRENT FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: PCT/EP99/03895
; PRIOR FILING DATE: 1999-06-04
; PRIOR APPLICATION NUMBER: 09/092,029
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 09/314,967
; PRIOR FILING DATE: 1999-05-20
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1042
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-959-392-2

Query Match 14.8%; Score 692; DB 4; Length 1042;
Best Local Similarity 27.8%; Pred. No. 3.8e-42;
Matches 210; Conservative 98; Mismatches 252; Indels 196; Gaps 35;
QY 202 DSKTVQRTQDNCSCFGLHARGVELMRTTPGFPDPSVPYPAHARQWALRGDAD----- 253
Db 365 DHDCVDSKDEVNCS--CHSQGLVECR-----NQCIPTSTFQCD-----GDECDKGSDEE 412
QY 254 --SVLSLTFRSFD-----LASCDERGSDLVTVYNTL---SPMEPHAVQLCGTYPSP 300
Db 413 NCSVIQTSCEGDQRCLYNPCLDSCG--GSSLCDPNNSLNCSQCEPITL--ELCWNLP-- 467
QY 301 YNLTFHSSQNVLLITLITNTERRHPGFETFFQLPRSSCCGRLRK-AQGTFFNPPYPGH 359

Db 468 YNST-----SYPNY-----FGHRTQKEASISWESSLPAL 497
QY 360 YPNIDICTWNIEVNNQHVKSFKFYLLPEPGVAGTCPKDYVEIN-GEK-----Y 409
Db 498 VQTN-----CYKLMFPFSCITLVPKC-----DVTGERIPPCRALCEH 535
QY 410 CGERSQFVV-----TSNKNKITVRPHSDQSYDTGFLA-EYLSYDSDSPC-PGOF 457
Db 536 SKERCSVLGIVGLWPEDTDCS-----QPPEENSQNTCLMPDEVV-----ECCSPSHF 585
QY 458 TCHTGCIRKELCDGWADCTDHSDELNCSCDAGHFTC-KNFKCPLFWVCDSDVNDG 516
Db 586 KCSGOCVLASRCDGADDDSDDEENCGCKERDLWECPSNKCQCLKHTVTCDFPDCPD 645
QY 517 NSDEQCS-CPAQTFRCNGKLSKQOQNGKDDCGDGDSEACPKVNV-----564
Db 646 YNDEKNCSPDDELECANHACVSRDLWCDGEADCSDSDEWDVTLINNVSSSFLMVH 705
QY 565 -----VTC-----TKHT-YRLNGLCL 580
Db 706 RAATEHHVCADGQOEILSQLACKQMGLEBPSVTKLIOEQEKPRWLTLSHNWESLNGTTL 765
QY 581 SK---GNPCDGEDSDGSDXDCDGLRSFTR-QARVVGGTDADEGEWPMQVSLHALG 636
Db 766 HELLVNGQCESKSKISLLCTKQ--DCGRRPAARMNKRILGRTSRPGRWPMQCSLQSRP 823
QY 637 QGHICGASLISPNWLVSAAHCYIDDRGFYSDPTQWTAFLGLHDQSORAPGVOERLKR 696
Db 824 SGHICCVLIACKWLTVAHCF-EGR-----ENNAWVKVGLGINNLDHPSV-FMQTRFVKT 877
QY 697 IISHPPFNDFPDYDIALLELEKPAEYSSMVRPICLPDASHVPPAGKAIWVTGWGHTQYG 756
Db 878 IILHPRYSRAVVDYDISIVELSDSETGVVRVCLPNPQWLEPDTYCVITGWGH--MG 935
QY 757 GTGALLIQGEIRVINTTCENLLPQO-ITPRMVCVGLSGVDSQCGSGGLSSVEAD 815
Db 936 NKMPFKLQGEVRILSLEHCQSYFDMKTTITRMICAGYESGTVDSQMGSDGGLVCEKPG 995
QY 816 GRIFOAGVSWGSGCAQR-NKPGVYTRLPFLFRDWIK 850
Db 996 GRWTLFGLTWSGSCVFSKVLGPGVYSNVSYFVEWIK 1031

RESULT 11
US-09-959-392-4
; Sequence 4, Application US/09959392
; Patent No. 6806075
; GENERAL INFORMATION:
; APPLICANT: MORSE, MICHAEL JOHN
; APPLICANT: WU, QINGYU
; APPLICANT: YAN, WEI
; TITLE OF INVENTION: CORIN, A SERINE PROTEASE
; FILE REFERENCE: BERLX 74A
; CURRENT APPLICATION NUMBER: US/09/959,392
; CURRENT FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: PCT/EP99/03895
; PRIOR FILING DATE: 1999-06-04
; PRIOR APPLICATION NUMBER: 09/092,029
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 09/314,967
; PRIOR FILING DATE: 1999-05-20
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 1113
; TYPE: PRT
; ORGANISM: Mus sp.
US-09-959-392-4

Query Match 14.2%; Score 664.5; DB 4; Length 1113;
Best Local Similarity 29.8%; Pred. No. 4.2e-40;
Matches 176; Conservative 78; Mismatches 221; Indels 115; Gaps 22;

QY 352 NSPYTPGHYP-----PNIDCTWNIE-----VNNQHVKSFKFYLLPEPGVAGTCP 398
Db 533 NLPYNHTHPYNYLGHRTQKEASISWESSLPALVQNTCYKLMFPFACITLVPKCDVNT-- 590
QY 399 KDYVEINGEK-----YCGERSQFVV-----TSNKNKITVRPHSDQSYDTGTF 440
Db 591 -----GQRIPPCRLLCESKXCESVLGIVGLWPEDTDCN-----QPPEESSDQNTCL 639
QY 441 LAEYLSYDSSDPCPQOCTCRGCIKELRCDCGMADCTDHSDELNCSCDAGHFTCK-NK 499
Db 640 LP---NEDEVESPSHFKCRSGRCVILGSRCDGQADDDSDDEENCGCKERLWECFPNK 696
QY 500 FCKPLFWVCDSDVNDGDNSEQCS-CPAQTFRCNGKLSKQOQNGKDDCGDGDSEAS 558
Db 697 QCLKHTLICDGPDPDPSMDENKCSFCQDNELECANHECVPRDLWCDGWVDCSDSDWEG 756
QY 559 CPKVN-----VVTCTKHTYRLNG-----LC-----LSKGNPECDGKE 591
Db 757 CVTLKNGNSSSLTIVHKSAREHVCADGWRETLSQLACKQMGLEBPSVTKLIPGQEQQ 816
QY 592 -----DCSDGSDKED-----C---DCGLRSFTR-QARVVGGTDD 622
Db 817 WURLYPNNENLNGSTLQELLVYRHSCPSRSEISLLCSKQDCGRRPAARMNKRILGRTSR 876
QY 623 EGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYIDDRGFYSDPTQWTAFLGLHDQ 682
Db 877 PGRWPMQCSLQSEPSGHICGCVLIACKWLTVAHCF-EGR-----EDADVWKVVGINNLD 931
QY 683 QRSAPGVOERLKRILSHPPFNDFPDYDIALLELEKPAEYSSMVRPICLPDASHVFPAG 742
Db 932 HPSG-FMQTRFVKTILLHPRYSRAVVDYDISIVELSDINETSYYVRVCLPSPPEEYLPD 990
QY 743 KAIWVTGWGHTQYGGTGALILQKGEIRVINTTCENLLPQO-ITPRMVCVGLSGVDS 801
Db 991 TYCITGWGH--MGNMFPKLOEGEVRIIPLEQCSYFDMKTTITRMICAGYESGTVDS 1048
QY 802 QGDSGGLSSVEADGRIFQAGVSWGSGCAQR-NKPGVYTRLPFLFRDWIK 850
Db 1049 MGDSGGLVCPERPQGWTLFGLTWSGSCVFSKVLGPGVYSNVSYFVGWIE 1098

RESULT 12

US-09-734-675-4
; Sequence 4, Application US/09734675
; Patent No. 6365391
; GENERAL INFORMATION:
; APPLICANT: WEBSTER, Marion et al
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL000862
; CURRENT APPLICATION NUMBER: US/09/734,675
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 407
; TYPE: PRT
; ORGANISM: Human
US-09-734-675-4

Query Match 12.6%; Score 588; DB 3; Length 407;
Best Local Similarity 45.9%; Pred. No. 4.7e-35;
Matches 119; Conservative 38; Mismatches 82; Indels 20; Gaps 6;

QY 604 CGLR---SFTROARVVGTDADGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYTD 660
Db 162 CGTRRSKTLQSLRIVGGTEVEGEWPMQASLQWDG-SHRCGATLINATLWLSAAHCT- 219
QY 661 DRGFYSDPTQWTAFLGLHDQSORAPGVOERLKRILSHPPFNDFPDYDIALLELEK 720
Db 220 ----TYKNPARWTASFGV-----TIKPSMKRGLRRIIVHEKYPKSHDYDISLAELSSP 270

QY 721 AYSMSVRPCLPDASHVFPAGKAIWVTGHTQYGGTGALILQKGEIRVINOTTCENLL 780
Db 271 VPTNAVHRVCLPDASYEFQPGDMFVTGALKNDGYSQHLRQAQVTLIDATTCE-- 328
QY 781 PQ----QITPRMVCVGLSGVSDCGSGPLSSVEADGRIFQAGVVSWGDCACARNKP 836
Db 329 PQAYNDAITPRMVCAGLEGTACQSGPLVSSDARDIWIYLAGIVSWGDCACARNKP 388
QY 837 GVTYTRPLFRDMIXENTGV 855
Db 389 GVTYTRVALRDMITSKTI 407

RESULT 13
US-09-949-016-11182
; Sequence 11182, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11182
; LENGTH: 717
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-11182

Query Match 12.3%; Score 577; DB 4; Length 717;
Best Local Similarity 25.3%; Pred. No. 6.2e-34;
Matches 193; Conservative 110; Mismatches 269; Indels 192; Gaps 31;

QY 203 SKTVQRTQDNCSCFGLHARGVELMRFTTPGPPDSPYPAHARCOWALRGDADSVLSITFRS 262
Db 33 SKASHTVELNMFG-----QIQSPGYPDS-YPSDSEVTWNITVPDGRIFKLYFMH 82
QY 263 FDLASCDERGSDLVTVNTLSPMEHALVQLCG-----TTPPSY-NLTFH 306
Db 83 FNLESSYLCEYDYVKV-----ETEDQVLATFCGRETDTTQTPGQEVVLSPGSFMISITPR 137
QY 307 SSQNVLLITLITNTERRHPGFEATFFQLP-----ETEDQVLATFCGRETDTTQTPGQEVVLSPGSFMISITPR 335
Db 138 SD-----FSENERFTGDAHYMAVDVDECKERDEBELSCDHYCHNYIGGYCSCRF 188
QY 336 -----RMSGCGRLRKAQGTNSPYPGHYPPNIDCTWNIENPNNHVKVSPKF 384
Db 189 GYLHTDNRTCRVCECDNLFTQRTGVTSPDPFNPYKSSSECLYTIIELEGFWNLQPED 248
QY 385 FYLLE--PGVPAGTCKPKDYVEIN-GEK-----YCGERSQFVVTNSNKTIVRPHSDQSYTD 437
Db 249 IFDIEDHPEVP--CPDYIKIKVGPVKVLGPPFCGEKAPBIPISQSHSVLILFHSNDSNGEN 305
QY 438 TGFLAELYSVDSDDPCP-----GQFTCRGTCIRKELRCDCGWADCTDHSDELNCSCDA 490
Db 306 RGWRLSYRA--AGNECEPELOPPVHGKLEPSQAKYFFK-----DQVLVSCDT 349
QY 491 GHQF-----TCRKNFKCLPLW-----VCSVNDGNSDEQGCSCPAQ-----TFR 531
Db 350 GYKVLKDNVEMDFQIECLDKGTWSNKIPTCKIVD-----CRAPGELEHGLITS 399
QY 532 CSNGKCLSSQCGKXGDCGSDGDEACPKVNVVCTKITYRC-----LNGLCISKGNP 585
Db 400 TRNNLTYYKSEI---KYSQOE-----PYKMLNNTGIYTCSAQGVMMNKV-LGRSLP 448

QY 586 ECDGKEDCSDGSDKDCDCGLRSFTRO--ARVVGTTDADEGEWPMQVSLHALCOQHICGA 643
Db 449 TCLPV-----CGLPKFSKLMARFNGRPAQKGTTPWTAMLSHLNGQFPFCGG 495
QY 644 SLISPNWLVSAAHCY-----IDDRFRYSD---PTQWTAFLGLHDOSQSAPOVQERRLK 695
Db 496 SLUGSSWIVTAACHLQSLDPEDPTLRSDLLSPSDFKIILGKH-WRLASDENEQHLGVK 554
QY 696 RIISHPFENDFDYDIALLELEKPAEYSSMVRPCLPDASHVFPAGKAIWVTGHTQY 755
Db 555 HTTLHPQYDPNTFENDVALVELLESPVNAFWNPICLPEGPO--QEGAMVIVSGWG-KQF 611
QY 756 GGTGALILQKGEIRVINOTTCENL---LPQQITPRMVCVGLSGVSDCGSGPLSSV 812
Db 612 LQRPPELTMEIPIVDHSTCQKAVAPLKCKVTRDMICAGEKGGKDACAGDSGGGPMVTL 671
QY 813 EAD-GRIFQAGVVSWGDCGCAQBNKPGVYTRLPFRDNIKENTGV 855
Db 672 NRERGQYLVGTVSWGDDCGKRDYGVSYIHNNKDWIQRVTGV 715

RESULT 14
US-09-949-016-6138
; Sequence 6138, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6138
; LENGTH: 699
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-6138

Query Match 12.3%; Score 574; DB 4; Length 699;
Best Local Similarity 25.3%; Pred. No. 1e-33;
Matches 193; Conservative 111; Mismatches 268; Indels 192; Gaps 32;

QY 203 SKTVQRTQDNCSCFGLHARGVELMRFTTPGPPDSPYPAHARCOWALRGDADSVLSITFRS 262
Db 15 SKASHTVELNMFG-----QIQSPGYPDS-YPSDSEVTWNITVPDGRIFKLYFMH 64
QY 263 FDLASCDERGSDLVTVNTLSPMEHALVQLCG-----TTPPSY-NLTFH 306
Db 65 FNLESSYLCEYDYVKV-----ETEDQVLATFCGRETDTTQTPGQEVVLSPGSFMISITPR 119
QY 307 SSQNVLLITLITNTERRHPGFEATFFQLP-----ETEDQVLATFCGRETDTTQTPGQEVVLSPGSFMISITPR 335
Db 120 SD-----FSENERFTGDAHYMAVDVDECKERDEBELSCDHYCHNYIGGYCSCRF 170
QY 336 -----RMSGCGRLRKAQGTNSPYPGHYPPNIDCTWNIENPNNHVKVSPKF 384
Db 171 GYLHTDNRTCRVCECDNLFTQRTGVTSPDPFNPYKSSSECLYTIIELEGFWNLQPED 230
QY 385 FYLLE--PGVPAGTCKPKDYVEIN-GEK-----YCGERSQFVVTNSNKTIVRPHSDQSYTD 437
Db 231 IFDIQDHPEVP--CPDYIKIKVGPVKVLGPPFCGEKAPBIPISQSHSVLILFHSNDSN 287
QY 438 TGFLAELYSVDSDDPCP-----GQFTCRGTCIRKELRCDCGWADCTDHSDELNCSCDA 490

Db 288 RGRSLSYRA--AGNECEPQPPVHGKIEPSQAKYFFK-----DQVLVSCDT 331
Qy 491 GHQF-----TCKKFKCPLFW-----VCSDVNDGSDNSDEQGCSPAQ-----TFR 531
Db 332 GYKVLKDNVEMDTFOIECLKDGTSNKIPTCKIVD-----CRAPGELEHGLITFS 381
Qy 532 CNGKCLSKSQOQNGKDDCGSDSDEASCPKVNVTCTKHYRC-----LNGLCCLSKGNP 585
Db 382 TRNNLTYYKSEI---KYSQOE-----PYKMLNNTGIYTCSAQGVMMKV-LGRSLP 430
Qy 586 ECDGKEDCSDGSDKDCGLRSFTQ--ARVVGTTDADEGEWPMQVSLHALGQGHICGA 643
Db 431 TCLPV-----CGLPKFSKULMARFNPRQKGTTPWTAMLSHLNGOPFCGG 477
Qy 644 SLISPNLVSAAHY---ID--DRGFRYS--PTOWTAFGLHDQSORSAPOVERRLK 695
Db 478 SILGSSWITAAHCLHSLDQDPTLRDSDLLSPDFKILLKH-WRLSDENEQHLGVK 536
Qy 696 RIISHFFNDFTDYDIALLELEKPAEYSSMRVPCLPDASHVFPAGKAIWVTGHTQY 755
Db 537 HTLHPQYDPTNFENDVALVELLESPLNAFVMPICLPEGPQ--QEGAMVIVSGWG-KQF 593
Qy 756 GTGALILQGEIRVINQTTCNL---LPQOITPRMCMVGLSGVDSCQDSCGSLSSV 812
Db 594 LQRPETLMEIPIIVDHTCQKAYAPLKKKVTDRDMICAGEKEGGKACAGSDGPGMVTL 653
Qy 813 EAD-GRIFOAGVYVSWGDCAQRNKPQVYTRPLFRDWMKENTGV 855
Db 654 NREGQWYLVGVYVSWGDDCGKDKORYGVSYIHHNKDWIORVTGV 697

RESULT 15

US-09-685-166A-895
; Sequence 895, Application US/09685166A
; Patent No. 6630305

GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqi
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Mark W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skelky, Yasir A.W.
; APPLICANT: Hepier, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C21
; CURRENT APPLICATION NUMBER: US/09/685,166A
; CURRENT FILING DATE: 2000-10-10
; NUMBER OF SEQ ID NOS: 898
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 895
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-685-166A-895

Query Match 12.0%; Score 560.5; DB 4; Length 492;
Best Local Similarity 33.9%; Pred. No. 6.2e-33;
Matches 131; Conservative 58; Mismatches 128; Indels 69; Gaps 14;

Qy 522 GCSCPAQTPRC--SNKCLSKSQOQNGKDDCGSDSDEASCPKV---NVV----- 565
Db 110 GSKCSNSGTECDSSGTCINPSNWCDBVSHCPGGEENRCVLYGPNFIQMYSSQKSKWH 169

Qy 566 -----TCT-----KHYRCLNG-----LCLSKGNPECDGKEDCSD 595
Db 170 PVCQDDWNNENYGRAACRDMGYKNNFYSSOGIVDDSGSTSPMKLNTSAGNVDIYKKLYHSD 229
Qy 596 GSDEK-----DCDCGLRSFTQARVVGTTDADEGEWPMQVSLHALGQGHICGASLIS 647
Db 230 ACSSKAVVSLRCLACGVNLNS-SRQSRIVGGESALPGAWPMQVSLH-VQNVHVCGGSIIT 287
Qy 648 PNWLVSAAHCYIDDRGFRYSPTOWTAFGLHDQS-QRSAPGVQERRLKRILISHPPFNDP 706
Db 288 FEWIVTAAHCVEKP-----LNNPWHWTAFAGLRQSFMYGAGYQ---VQKVISHPNYDSK 340
Qy 707 TFDYDIALLELEKPAEYSSMRVPCLPDASHVFPAGKAIWVTGHTQYGGTGCALILQKG 766
Db 341 TKNDIALMKLOKPLTFNDLVKPCLPNPGMMLQPEQLCWISGWGATEBKGTSEVLNAA 400
Qy 767 EIRVINQTTCN--LLPQOITPRMCMVGLSGVDSCQDSCGSLSSVEADGRI-FOAGV 823
Db 401 KVLITETQRCNSRYVYDNLITPAMICAGFLQGNVDSQCGDSGGPL--VTSNNNIWWLIGD 458
Qy 824 VSWGDCGAQRNKPQVYTRPLFRDWM 849
Db 459 TSWGSGCAKAYRPGVYGNVMVFTDWI 484

Search completed: September 23, 2005, 12:59:40

Job time : 29 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 23, 2005, 12:56:11 ; Search time 174 Seconds
(without alignments)
1999.974 Million cell updates/sec

Title: US-09-421-213-2
Perfect score: 4681
Sequence: 1 MGSDRARKGGGPKDFGAGL.....PGVYTRFLPRDWIKENTGV 855

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1826521 seqs, 407012169 residues

Total number of hits satisfying chosen parameters: 1826521

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :				Published Applications AA:**
1:	/cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*			
2:	/cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*			
3:	/cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*			
4:	/cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*			
5:	/cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*			
6:	/cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*			
7:	/cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*			
8:	/cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*			
9:	/cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*			
10:	/cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*			
11:	/cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*			
12:	/cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*			
13:	/cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*			
14:	/cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*			
15:	/cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*			
16:	/cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*			
17:	/cgn2_6/ptodata/1/pubpaa/US10E_PUBCOMB.pep.*			
18:	/cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*			
19:	/cgn2_6/ptodata/1/pubpaa/US11A_PUBCOMB.pep.*			
20:	/cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*			
21:	/cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*			
22:	/cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*			

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4681	100.0	855	10	US-09-776-191-2
2	4681	100.0	855	14	US-10-099-700A-2
3	4681	100.0	855	14	US-10-190-030B-2
4	4681	100.0	855	14	US-10-302-840A-2
5	4681	100.0	855	14	US-10-267-219-2
6	4681	100.0	855	14	US-10-112-221A-2
7	4681	100.0	855	14	US-10-144-271-2
8	4681	100.0	855	15	US-10-147-211A-2
9	4681	100.0	855	15	US-10-156-214A-2
10	4681	100.0	855	15	US-10-072-012-352
11	4681	100.0	855	15	US-10-072-012-411

12	4681	100.0	855	15	US-10-072-012-418
13	4681	100.0	855	15	US-10-600-187-2
14	4681	100.0	855	17	US-10-612-466B-2
15	4676	99.9	855	15	US-10-295-027-1185
16	4676	99.9	855	15	US-10-072-012-353
17	4676	99.9	855	15	US-10-072-012-412
18	4676	99.9	855	15	US-10-072-012-419
19	4672	99.8	855	15	US-10-072-012-354
20	4672	99.8	855	15	US-10-072-012-420
21	4672	99.8	855	15	US-10-037-417-132
22	4631	98.9	851	15	US-10-276-774-1798
23	4631	98.9	851	15	US-10-296-115-1143
24	4319	92.3	932	18	US-10-450-763-51801
25	4175.5	89.2	782	14	US-10-097-340-312
26	4175	89.2	762	16	US-10-729-807-1
27	4111	87.8	757	15	US-10-072-012-44
28	3901	83.3	855	9	US-09-900-751-2
29	3901	83.3	855	15	US-10-072-012-355
30	3901	83.3	855	15	US-10-072-012-413
31	3883	83.0	855	15	US-10-072-012-356
32	3883	83.0	855	15	US-10-072-012-414
33	3883	83.0	855	15	US-10-072-012-417
34	3810	81.4	902	15	US-10-333-743-3
35	3810	81.4	902	15	US-10-600-187-10
36	3810	81.4	902	16	US-10-297-987B-11
37	2980	63.7	620	9	US-09-925-301-1193
38	2664	56.9	845	15	US-10-072-012-415
39	1352	28.9	362	18	US-10-450-763-51800
40	1319	28.2	241	10	US-09-776-191-50
41	1319	28.2	241	14	US-10-099-700A-4
42	1319	28.2	241	14	US-10-092-004A-2
43	1319	28.2	241	14	US-10-190-030B-4
44	1319	28.2	241	14	US-10-302-840A-4
45	1319	28.2	241	14	US-10-267-219-4

ALIGNMENTS

RESULT 1

US-09-776-191-2
; Sequence 2, Application US/09776191
; Publication No. US20030119168A1
; GENERAL INFORMATION:
; APPLICANT: Edwin L. Madison
; APPLICANT: Edgar O. Ong
; APPLICANT: Jiunn-Chern Yeh
; APPLICANT: Corvas International, Inc.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
; TITLE OF INVENTION: TRANSMEMBRANE SERINE PROTEASES, THE ENCODED PROTEINS AND
; TITLE OF INVENTION: METHODS BASED THEREON
; FILE REFERENCE: 24745-1607
; CURRENT APPLICATION NUMBER: US/09/776,191
; CURRENT FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/213,124
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 60/234,840
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 60/179,982
; PRIOR FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: 60/183,542
; PRIOR FILING DATE: 2000-02-18
; PRIOR APPLICATION NUMBER: 09/657,968
; PRIOR FILING DATE: 2000-02-08
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-776-191-2

Query Match 100.0%; Score 4681; DB 10; Length 855;

Best Local Similarity 100.0%; Pred. No. 0; Matches 855; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	1	MGSDRARKGGGPKDFGAGLKYNRHEKVNGLKEEGVEFLPVNNVKKVKGHGPRWVLLAA	60						
Db	1	MGSDRARKGGGPKDFGAGLKYNRHEKVNGLKEEGVEFLPVNNVKKVKGHGPRWVLLAA	60						
QY	61	VLIGLLLVLLGIGFVLVHQLQYRDVRVQKVFNGYMRITNENFVDAYENSNSTEFVSLASKV	120						
Db	61	VLIGLLLVLLGIGFVLVHQLQYRDVRVQKVFNGYMRITNENFVDAYENSNSTEFVSLASKV	120						
QY	121	KDALKLLYSGVPFLGPYHKESAVTAFSEGSVIAYYWSFSPQHLYVEAEARVMAEERVVM	180						
Db	121	KDALKLLYSGVPFLGPYHKESAVTAFSEGSVIAYYWSFSPQHLYVEAEARVMAEERVVM	180						
QY	181	LPPEARSLKSFVVTSVVAFPTDSKTQRTQDNSCSFGLHARGVELMRTTTPGFPDPSYPYA	240						
Db	181	LPPEARSLKSFVVTSVVAFPTDSKTQRTQDNSCSFGLHARGVELMRTTTPGFPDPSYPYA	240						
QY	241	HARQWALRGDADSVLSLTFRSFDLASCDERGSDLVTVYNTLSPMEPHALVOLCGTTPPS	300						
Db	241	HARQWALRGDADSVLSLTFRSFDLASCDERGSDLVTVYNTLSPMEPHALVOLCGTTPPS	300						
QY	301	YNLTFHSSQNVLLITLITNTERRHPGFEATFFQLPRMSSCGGLRKAQGTFTNSPYPGHY	360						
Db	301	YNLTFHSSQNVLLITLITNTERRHPGFEATFFQLPRMSSCGGLRKAQGTFTNSPYPGHY	360						
QY	361	PNNICTWNIIEVPNNQHVSKFYYLLEPGVPAGTCKPDYVEINGEKYGERSQFVVT	420						
Db	361	PNNICTWNIIEVPNNQHVSKFYYLLEPGVPAGTCKPDYVEINGEKYGERSQFVVT	420						
QY	421	NSNKITVRPHSDQSYTDTGFLAEYLSYDSSDPCPGQFTCTGRCIRKELRCDGMADCTDH	480						
Db	421	NSNKITVRPHSDQSYTDTGFLAEYLSYDSSDPCPGQFTCTGRCIRKELRCDGMADCTDH	480						
QY	481	SDELNCSCDAGHQFTCKNFKCPKLVFWVCDSDVNDGDNDEQSCCPAQTFRCNSGKCLSK	540						
Db	481	SDELNCSCDAGHQFTCKNFKCPKLVFWVCDSDVNDGDNDEQSCCPAQTFRCNSGKCLSK	540						
QY	541	SOQNGKDDCGSDGSDASCPKVVVTCTKTYRCLNGLCLSKGNPECDGKEDCSGSDK	600						
Db	541	SOQNGKDDCGSDGSDASCPKVVVTCTKTYRCLNGLCLSKGNPECDGKEDCSGSDK	600						
QY	601	DCDCGLRSFTRQARVVGTDADGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYID	660						
Db	601	DCDCGLRSFTRQARVVGTDADGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYID	660						
QY	661	DRGFYSDPTQWTAFLGLHDQORSAPGVQERRLKRRIISHPPFNDFTFDYDIALLELEKP	720						
Db	661	DRGFYSDPTQWTAFLGLHDQORSAPGVQERRLKRRIISHPPFNDFTFDYDIALLELEKP	720						
QY	721	ABYSSMVRPILCLPDASHVFPAGKAIWVTGHWGHTQYGGTGALLQKGEIRVINQTCENLL	780						
Db	721	ABYSSMVRPILCLPDASHVFPAGKAIWVTGHWGHTQYGGTGALLQKGEIRVINQTCENLL	780						
QY	781	POQITPRMVCVGLSGGVDSQCGSDGGLSSVEADGRIFQAGVSWGDCGAQRNKPQVYT	840						
Db	781	POQITPRMVCVGLSGGVDSQCGSDGGLSSVEADGRIFQAGVSWGDCGAQRNKPQVYT	840						
QY	841	RLPLFRDWIKENTGV	855						
Db	841	RLPLFRDWIKENTGV	855						

RESULT 2
US-10-099-700A-2
; Sequence 2, Application US/10099700A
; Publication NO. US2003008372A1
; GENERAL INFORMATION:
; APPLICANT: Edwin L. Madison
; APPLICANT: Edgar O. Ong
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING TRANSMEMBRANE SERINE PROTEASE 7,
; TITLE OF INVENTION: ENCODED POLYPEPTIDES AND METHODS BASED THEREON

; FILE REFERENCE: 24745-1613
; CURRENT APPLICATION NUMBER: US/10/099,700A
; CURRENT FILING DATE: 2002-05-24
; PRIOR APPLICATION NUMBER: 60/275,592
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo Sapien
; US-10-099-700A-2

Query Match 100.0%; Score 4681; DB 14; Length 855; Best Local Similarity 100.0%; Pred. No. 0; Matches 855; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	1	MGSDRARKGGGPKDFGAGLKYNRHEKVNGLKEEGVEFLPVNNVKKVKGHGPRWVLLAA	60						
Db	1	MGSDRARKGGGPKDFGAGLKYNRHEKVNGLKEEGVEFLPVNNVKKVKGHGPRWVLLAA	60						
QY	61	VLIGLLLVLLGIGFVLVHQLQYRDVRVQKVFNGYMRITNENFVDAYENSNSTEFVSLASKV	120						
Db	61	VLIGLLLVLLGIGFVLVHQLQYRDVRVQKVFNGYMRITNENFVDAYENSNSTEFVSLASKV	120						
QY	121	KDALKLLYSGVPFLGPYHKESAVTAFSEGSVIAYYWSFSPQHLYVEAEARVMAEERVVM	180						
Db	121	KDALKLLYSGVPFLGPYHKESAVTAFSEGSVIAYYWSFSPQHLYVEAEARVMAEERVVM	180						
QY	181	LPPEARSLKSFVVTSVVAFPTDSKTQRTQDNSCSFGLHARGVELMRTTTPGFPDPSYPYA	240						
Db	181	LPPEARSLKSFVVTSVVAFPTDSKTQRTQDNSCSFGLHARGVELMRTTTPGFPDPSYPYA	240						
QY	241	HARQWALRGDADSVLSLTFRSFDLASCDERGSDLVTVYNTLSPMEPHALVOLCGTTPPS	300						
Db	241	HARQWALRGDADSVLSLTFRSFDLASCDERGSDLVTVYNTLSPMEPHALVOLCGTTPPS	300						
QY	301	YNLTFHSSQNVLLITLITNTERRHPGFEATFFQLPRMSSCGGLRKAQGTFTNSPYPGHY	360						
Db	301	YNLTFHSSQNVLLITLITNTERRHPGFEATFFQLPRMSSCGGLRKAQGTFTNSPYPGHY	360						
QY	361	PNNICTWNIIEVPNNQHVSKFYYLLEPGVPAGTCKPDYVEINGEKYGERSQFVVT	420						
Db	361	PNNICTWNIIEVPNNQHVSKFYYLLEPGVPAGTCKPDYVEINGEKYGERSQFVVT	420						
QY	421	NSNKITVRPHSDQSYTDTGFLAEYLSYDSSDPCPGQFTCTGRCIRKELRCDGMADCTDH	480						
Db	421	NSNKITVRPHSDQSYTDTGFLAEYLSYDSSDPCPGQFTCTGRCIRKELRCDGMADCTDH	480						
QY	481	SDELNCSCDAGHQFTCKNFKCPKLVFWVCDSDVNDGDNDEQSCCPAQTFRCNSGKCLSK	540						
Db	481	SDELNCSCDAGHQFTCKNFKCPKLVFWVCDSDVNDGDNDEQSCCPAQTFRCNSGKCLSK	540						
QY	541	SOQNGKDDCGSDGSDASCPKVVVTCTKTYRCLNGLCLSKGNPECDGKEDCSGSDK	600						
Db	541	SOQNGKDDCGSDGSDASCPKVVVTCTKTYRCLNGLCLSKGNPECDGKEDCSGSDK	600						
QY	601	DCDCGLRSFTRQARVVGTDADGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYID	660						
Db	601	DCDCGLRSFTRQARVVGTDADGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYID	660						
QY	661	DRGFYSDPTQWTAFLGLHDQORSAPGVQERRLKRRIISHPPFNDFTFDYDIALLELEKP	720						
Db	661	DRGFYSDPTQWTAFLGLHDQORSAPGVQERRLKRRIISHPPFNDFTFDYDIALLELEKP	720						
QY	721	ABYSSMVRPILCLPDASHVFPAGKAIWVTGHWGHTQYGGTGALLQKGEIRVINQTCENLL	780						
Db	721	ABYSSMVRPILCLPDASHVFPAGKAIWVTGHWGHTQYGGTGALLQKGEIRVINQTCENLL	780						
QY	781	POQITPRMVCVGLSGGVDSQCGSDGGLSSVEADGRIFQAGVSWGDCGAQRNKPQVYT	840						
Db	781	POQITPRMVCVGLSGGVDSQCGSDGGLSSVEADGRIFQAGVSWGDCGAQRNKPQVYT	840						
QY	841	RLPLFRDWIKENTGV	855						
Db	841	RLPLFRDWIKENTGV	855						

```
QY      841 RLPLFRDWIKENTGV 855
Db      841 RLPLFRDWIKENTGV 855

RESULT 3
US-10-190-030B-2
; Sequence 2, Application US/10190030B
; Publication No. US20030134298A1
; GENERAL INFORMATION:
; APPLICANT: Madison, Edwin
; APPLICANT: Ong, Edgar
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING A
; TITLE OF INVENTION: TRANSMEMBRANE SERINE PROTEASE 20, THE ENCODED POLYPEPTIDES AND
; TITLE OF INVENTION: METHODS BASED THEREON
; FILE REFERENCE: 24745-1618
; CURRENT APPLICATION NUMBER: US/10/190,030B
; CURRENT FILING DATE: 2002-07-03
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-190-030B-2

Query Match      100.0%; Score 4681; DB 14; Length 855;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 855; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MGSDRARKGGGPKDFGAGLKYSRHEKVGLEEGVEFLPVNNVKKVKGHGRWVLA 60
Db      1 MGSDRARKGGGPKDFGAGLKYSRHEKVGLEEGVEFLPVNNVKKVKGHGRWVLA 60

QY      61 VLIGLLVLLGIGFLVWHLOQYRDVVRQKVFNGYMRITNENFVDAYENSNSSTFVSLASKV 120
Db      61 VLIGLLVLLGIGFLVWHLOQYRDVVRQKVFNGYMRITNENFVDAYENSNSSTFVSLASKV 120

QY      121 KDALKLYSGVPLGPGYHKESAVTAFSEGSVIAYWSEFSIPQHLVEEAERVMAEERVVM 180
Db      121 KDALKLYSGVPLGPGYHKESAVTAFSEGSVIAYWSEFSIPQHLVEEAERVMAEERVVM 180

QY      181 LPPRARSLSKSFVWTSVVAFTDSTKVTQDNSCSFGLHARGVELMRFTTTCGFPDPSYP 240
Db      181 LPPRARSLSKSFVWTSVVAFTDSTKVTQDNSCSFGLHARGVELMRFTTTCGFPDPSYP 240

QY      241 HARCOWALRGDADSVLSLTFRSFDLASCDERSGDLVTVYNTLSPMEPHALVOLCGTYP 300
Db      241 HARCOWALRGDADSVLSLTFRSFDLASCDERSGDLVTVYNTLSPMEPHALVOLCGTYP 300

QY      301 YNLTFHSSQNVLLITLITNTERRHPGFEATFFQLPRMSSCGRLRKAQGTFSNYPY 360
Db      301 YNLTFHSSQNVLLITLITNTERRHPGFEATFFQLPRMSSCGRLRKAQGTFSNYPY 360

QY      361 PPNIDCTWNIIEVPNNQHVKSFKFYLLEPGVPAGTCPKDYVEINGEKYCGERSQFV 420
Db      361 PPNIDCTWNIIEVPNNQHVKSFKFYLLEPGVPAGTCPKDYVEINGEKYCGERSQFV 420

QY      421 NSNKITVRPHSDQSYDTDTGFLAEYLSYDSSDPCPGQFTCRGRCIRKELRCDCGWADCTDH 480
Db      421 NSNKITVRPHSDQSYDTDTGFLAEYLSYDSSDPCPGQFTCRGRCIRKELRCDCGWADCTDH 480

QY      481 SDELINCSCDAGHQFTCKNFKCPLFWVCDVNDGDNDEQSCPCPAQTFRCSNGKCLSK 540
Db      481 SDELINCSCDAGHQFTCKNFKCPLFWVCDVNDGDNDEQSCPCPAQTFRCSNGKCLSK 540

QY      541 SQQCNKGKDCGQSDSEASCPKVVVTCTKHTVRCNLGLCLSKGNPECDGKEDCSGDBK 600
Db      541 SQQCNKGKDCGQSDSEASCPKVVVTCTKHTVRCNLGLCLSKGNPECDGKEDCSGDBK 600

QY      601 DCDGGLSRFTQARVVGTTDADGEGWPQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
Db      601 DCDGGLSRFTQARVVGTTDADGEGWPQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
```

```
QY      661 DRGFYSDPTQWTAFLGLHDQSORGAPGVQERRLKRILSHSHPFNFDFDYDIALLEK 720
Db      661 DRGFYSDPTQWTAFLGLHDQSORGAPGVQERRLKRILSHSHPFNFDFDYDIALLEK 720

QY      721 ABYSNMVRPICLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTTCE 780
Db      721 ABYSNMVRPICLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTTCE 780

QY      781 POQITPRMCMVGFSLGGVDSQCGDGGPLSSVEADGRIFQAGVSWGDCGCAORNKPGV 840
Db      781 POQITPRMCMVGFSLGGVDSQCGDGGPLSSVEADGRIFQAGVSWGDCGCAORNKPGV 840

QY      841 RLPLFRDWIKENTGV 855
Db      841 RLPLFRDWIKENTGV 855

RESULT 4
US-10-302-840A-2
; Sequence 2, Application US/10302840A
; Publication No. US20030134794A1
; GENERAL INFORMATION:
; APPLICANT: Madison, Edwin L.
; APPLICANT: Ong, Edgar O.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING SERINE PROTEASE CVSP17, THE ENC
; TITLE OF INVENTION: POLYPEPTIDES AND METHODS BASED THEREON
; FILE REFERENCE: 24745-1622
; CURRENT APPLICATION NUMBER: US/10/302,840A
; CURRENT FILING DATE: 2003-01-24
; PRIOR APPLICATION NUMBER: 60/332,015
; PRIOR FILING DATE: 2001-11-20
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-302-840A-2

Query Match      100.0%; Score 4681; DB 14; Length 855;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 855; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MGSDRARKGGGPKDFGAGLKYSRHEKVGLEEGVEFLPVNNVKKVKGHGRWVLA 60
Db      1 MGSDRARKGGGPKDFGAGLKYSRHEKVGLEEGVEFLPVNNVKKVKGHGRWVLA 60

QY      61 VLIGLLVLLGIGFLVWHLOQYRDVVRQKVFNGYMRITNENFVDAYENSNSSTFVSLASKV 120
Db      61 VLIGLLVLLGIGFLVWHLOQYRDVVRQKVFNGYMRITNENFVDAYENSNSSTFVSLASKV 120

QY      121 KDALKLYSGVPLGPGYHKESAVTAFSEGSVIAYWSEFSIPQHLVEEAERVMAEERVVM 180
Db      121 KDALKLYSGVPLGPGYHKESAVTAFSEGSVIAYWSEFSIPQHLVEEAERVMAEERVVM 180

QY      181 LPPRARSLSKSFVWTSVVAFTDSTKVTQDNSCSFGLHARGVELMRFTTTCGFPDPSYP 240
Db      181 LPPRARSLSKSFVWTSVVAFTDSTKVTQDNSCSFGLHARGVELMRFTTTCGFPDPSYP 240

QY      241 HARCOWALRGDADSVLSLTFRSFDLASCDERSGDLVTVYNTLSPMEPHALVOLCGTYP 300
Db      241 HARCOWALRGDADSVLSLTFRSFDLASCDERSGDLVTVYNTLSPMEPHALVOLCGTYP 300

QY      301 YNLTFHSSQNVLLITLITNTERRHPGFEATFFQLPRMSSCGRLRKAQGTFSNYPY 360
Db      301 YNLTFHSSQNVLLITLITNTERRHPGFEATFFQLPRMSSCGRLRKAQGTFSNYPY 360

QY      361 PPNIDCTWNIIEVPNNQHVKSFKFYLLEPGVPAGTCPKDYVEINGEKYCGERSQFV 420
Db      361 PPNIDCTWNIIEVPNNQHVKSFKFYLLEPGVPAGTCPKDYVEINGEKYCGERSQFV 420

QY      421 NSNKITVRPHSDQSYDTDTGFLAEYLSYDSSDPCPGQFTCRGRCIRKELRCDCGWADCTDH 480
```

Db 421 NSNKITVRHSDQSYTDGFLAEYLSYDSSDPCGQFCTRTGRCIRKELRCGWDCTDH 480
QY 481 SDELNCSADAGHFTCKNFKCKPLFWVCDVNDGNSDEQGSCEPAQTFRCSNGKCLSK 540
Db 481 SDELNCSADAGHFTCKNFKCKPLFWVCDVNDGNSDEQGSCEPAQTFRCSNGKCLSK 540
QY 541 SQQCNKDDCGDSDGSDASCPKVNVTCTKHTYRCLNGLCLSKNPECDGKEDCSGSDSK 600
Db 541 SQQCNKDDCGDSDGSDASCPKVNVTCTKHTYRCLNGLCLSKNPECDGKEDCSGSDSK 600
QY 601 DDCGRLSFTQARVVGTTDADGEGWPQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
Db 601 DDCGRLSFTQARVVGTTDADGEGWPQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
QY 661 DRGRYSDPTQWTAFLGLHDQSORQAPGVQERRLKRIISHPPFNDFTFDYDIALLELEKP 720
Db 661 DRGRYSDPTQWTAFLGLHDQSORQAPGVQERRLKRIISHPPFNDFTFDYDIALLELEKP 720
QY 721 AYEYSSMRVPCLPDASHVFPAGKAIWVTGWHTQYGGTGALILQKGEIRVINQTCENLL 780
Db 721 AYEYSSMRVPCLPDASHVFPAGKAIWVTGWHTQYGGTGALILQKGEIRVINQTCENLL 780
QY 781 PQOITPRMCMVGLSGGVDSQCGDSSGGLSSVEADGRIFQAGVWSGDCGCAQRNKPQVYT 840
Db 781 PQOITPRMCMVGLSGGVDSQCGDSSGGLSSVEADGRIFQAGVWSGDCGCAQRNKPQVYT 840
QY 841 RLPLFRDWIKENTGV 855
Db 841 RLPLFRDWIKENTGV 855
RESULT 5
US-10-267-219-2
; Sequence 2, Application US/10267219
; Publication No. US20030143219A1
; GENERAL INFORMATION:
; APPLICANT: Madison, Edwin
; APPLICANT: Yeh, Jium-Chern
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING A TRANSMEMBRANE SERINE PROTEASE 2
; TITLE OF INVENTION: ENCODED POLYPEPTIDES AND METHODS BASED THEREON
; FILE REFERENCE: 24745-1521
; CURRENT APPLICATION NUMBER: US/10/267,219
; CURRENT FILING DATE: 2002-10-08
; PRIOR APPLICATION NUMBER: 60/328,530
; PRIOR FILING DATE: 09-OCT-2001
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-267-219-2
Query Match 100.0%; Score 4681; DB 14; Length 855;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 855; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MGSDRARKGGGPKDFGAGLKYSRHEKYNLEEGVEFLPVNNVKKVEKHGPRWVWVLA 60
Db 1 MGSDRARKGGGPKDFGAGLKYSRHEKYNLEEGVEFLPVNNVKKVEKHGPRWVWVLA 60
QY 61 VLIGLLLVLLGFLVHLYQYRDVVRQKVFNGVMYRITNFENFVDAYENSNSTFVSLASKV 120
Db 61 VLIGLLLVLLGFLVHLYQYRDVVRQKVFNGVMYRITNFENFVDAYENSNSTFVSLASKV 120
QY 121 KDALKLLYGVPLGPGYHKESAVTAFSEGSVIAYYWSFSPHQLVVEAERVAERVVM 180
Db 121 KDALKLLYGVPLGPGYHKESAVTAFSEGSVIAYYWSFSPHQLVVEAERVAERVVM 180
QY 181 LPPRARSLSKSVVTSVVAFTDSTKVQRTQDNCSFGLHARGVELMRTFTPGFPPSPYPA 240
Db 181 LPPRARSLSKSVVTSVVAFTDSTKVQRTQDNCSFGLHARGVELMRTFTPGFPPSPYPA 240

QY 241 HARCOWALRGDADSVLSLTFRSFDLASCDERGSDLVTVYNTLSPMEPHALVOLCGTTPS 300
Db 241 HARCOWALRGDADSVLSLTFRSFDLASCDERGSDLVTVYNTLSPMEPHALVOLCGTTPS 300
QY 301 YNLTFHSSQNVLLITLINTERRHPGFATFPOLPRMSSCGGRLRKAQOTFNSPYYPGHY 360
Db 301 YNLTFHSSQNVLLITLINTERRHPGFATFPOLPRMSSCGGRLRKAQOTFNSPYYPGHY 360
QY 361 PNINDCTWNIEVPNNQHVKSFKFYLLEPGVPAGTCKPKDYVEINGEKYCGERSQFVVT 420
Db 361 PNINDCTWNIEVPNNQHVKSFKFYLLEPGVPAGTCKPKDYVEINGEKYCGERSQFVVT 420
QY 421 NSNKITVRHSDQSYTDGFLAEYLSYDSSDPCGQFCTRTGRCIRKELRCGWDCTDH 480
Db 421 NSNKITVRHSDQSYTDGFLAEYLSYDSSDPCGQFCTRTGRCIRKELRCGWDCTDH 480
QY 481 SDELNCSADAGHFTCKNFKCKPLFWVCDVNDGNSDEQGSCEPAQTFRCSNGKCLSK 540
Db 481 SDELNCSADAGHFTCKNFKCKPLFWVCDVNDGNSDEQGSCEPAQTFRCSNGKCLSK 540
QY 541 SQQCNKDDCGDSDGSDASCPKVNVTCTKHTYRCLNGLCLSKNPECDGKEDCSGSDSK 600
Db 541 SQQCNKDDCGDSDGSDASCPKVNVTCTKHTYRCLNGLCLSKNPECDGKEDCSGSDSK 600
QY 601 DDCGRLSFTQARVVGTTDADGEGWPQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
Db 601 DDCGRLSFTQARVVGTTDADGEGWPQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
QY 661 DRGRYSDPTQWTAFLGLHDQSORQAPGVQERRLKRIISHPPFNDFTFDYDIALLELEKP 720
Db 661 DRGRYSDPTQWTAFLGLHDQSORQAPGVQERRLKRIISHPPFNDFTFDYDIALLELEKP 720
QY 721 AYEYSSMRVPCLPDASHVFPAGKAIWVTGWHTQYGGTGALILQKGEIRVINQTCENLL 780
Db 721 AYEYSSMRVPCLPDASHVFPAGKAIWVTGWHTQYGGTGALILQKGEIRVINQTCENLL 780
QY 781 PQOITPRMCMVGLSGGVDSQCGDSSGGLSSVEADGRIFQAGVWSGDCGCAQRNKPQVYT 840
Db 781 PQOITPRMCMVGLSGGVDSQCGDSSGGLSSVEADGRIFQAGVWSGDCGCAQRNKPQVYT 840
QY 841 RLPLFRDWIKENTGV 855
Db 841 RLPLFRDWIKENTGV 855
RESULT 6
US-10-112-221A-2
; Sequence 2, Application US/10112221A
; Publication No. US20030166851A1
; GENERAL INFORMATION:
; APPLICANT: Madison, Edwin
; APPLICANT: Ong, Edgar O.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING A TRANSMEMBRANE SERINE PROTEASE
; TITLE OF INVENTION: ENCODED POLYPEPTIDES AND METHODS BASED THEREON
; FILE REFERENCE: 24745-1615
; CURRENT APPLICATION NUMBER: US/10/112,221A
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: 60/279,228
; PRIOR FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: 60/291,501
; PRIOR FILING DATE: 2001-05-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-112-221A-2
Query Match 100.0%; Score 4681; DB 14; Length 855;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 855; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 MGS DRARKGGGPKDFGAGLKYNRHEKVNGLGEEVEFLPVNNVKVKEKHGGRWVLA 60
|
|
|
Db 1 MGS DRARKGGGPKDFGAGLKYNRHEKVNGLGEEVEFLPVNNVKVKEKHGGRWVLA 60
|
|
|
QY 61 VLIGLLVLLGIGFLVHLYQYRDVRVQKVFNGYMRITNENFVDAYENSNSTEFVSLASKV 120
|
|
|
Db 61 VLIGLLVLLGIGFLVHLYQYRDVRVQKVFNGYMRITNENFVDAYENSNSTEFVSLASKV 120
|
|
|
QY 121 KDALKLYSGVPFLGPHYKESAVTAFSEGSAVIAYWSEFSIPOHLVEEAERVAERVM 180
|
|
|
Db 121 KDALKLYSGVPFLGPHYKESAVTAFSEGSAVIAYWSEFSIPOHLVEEAERVAERVM 180
|
|
|
QY 181 LPPRARSLSKSFVVTSVVAFPTDSKTQRTQDNCSFGLHARGVELMRTTTPGFPDPSYP 240
|
|
|
Db 181 LPPRARSLSKSFVVTSVVAFPTDSKTQRTQDNCSFGLHARGVELMRTTTPGFPDPSYP 240
|
|
|
QY 241 HARCQWALRGDADSVLSLTFRSFDLASCDERSGDLVTYNTLSPMEPHALVOLCGTYP 300
|
|
|
Db 241 HARCQWALRGDADSVLSLTFRSFDLASCDERSGDLVTYNTLSPMEPHALVOLCGTYP 300
|
|
|
QY 301 YNLTFHSSQNVLLITLITNTERRHGPGFATFPQLPRMSSCGGLRKAQGTNSPYPGHY 360
|
|
|
Db 301 YNLTFHSSQNVLLITLITNTERRHGPGFATFPQLPRMSSCGGLRKAQGTNSPYPGHY 360
|
|
|
QY 361 PPNIDCTWNIIEVPNNQHVSKFVYLLPEPGVPAGTCKPKDYVEINGEKYCGERSQFV 420
|
|
|
Db 361 PPNIDCTWNIIEVPNNQHVSKFVYLLPEPGVPAGTCKPKDYVEINGEKYCGERSQFV 420
|
|
|
QY 421 NSNKITVRPHSDQSYTDTGFLAEYLSYDSSDPCPGQFTCTGRCIRKELRCDGWADCT 480
|
|
|
Db 421 NSNKITVRPHSDQSYTDTGFLAEYLSYDSSDPCPGQFTCTGRCIRKELRCDGWADCT 480
|
|
|
QY 481 SDELNCSCDAGHQFTCKNFKCKPLFWVCDSDVNDGDNDEQSCCPAQTFRCNSGKCLSK 540
|
|
|
Db 481 SDELNCSCDAGHQFTCKNFKCKPLFWVCDSDVNDGDNDEQSCCPAQTFRCNSGKCLSK 540
|
|
|
QY 541 SQQNGKDDCGDSDGDEASCPKVNVTCTKHTYRCLNGLCLSKGNPECDGKEDCDSDG 600
|
|
|
Db 541 SQQNGKDDCGDSDGDEASCPKVNVTCTKHTYRCLNGLCLSKGNPECDGKEDCDSDG 600
|
|
|
QY 601 DCDGLRSFTRQARVVGTDAGEWPMQVSLHALGQGHICGASLISPNWLVAACHYID 660
|
|
|
Db 601 DCDGLRSFTRQARVVGTDAGEWPMQVSLHALGQGHICGASLISPNWLVAACHYID 660
|
|
|
QY 661 DRGFRYSDPTQWTAFLGLHDQSORSAQGVQERRLRKRIISHPPFNDFTDYDIALLE 720
|
|
|
Db 661 DRGFRYSDPTQWTAFLGLHDQSORSAQGVQERRLRKRIISHPPFNDFTDYDIALLE 720
|
|
|
QY 721 AEYSWMVRPCLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTCENLL 780
|
|
|
Db 721 AEYSWMVRPCLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTCENLL 780
|
|
|
QY 781 PQQITPRMCMVGFSLGGVSDSCQDGGPLSSVEADGRIFQAGVSWGDCGAQRNKP 840
|
|
|
Db 781 PQQITPRMCMVGFSLGGVSDSCQDGGPLSSVEADGRIFQAGVSWGDCGAQRNKP 840
|
|
|
QY 841 RLPLFRDWIKENTGV 855
|
|
|
Db 841 RLPLFRDWIKENTGV 855
|
|
|
```

RESULT 7

```
US-10-104-271-2
; Sequence 2, Application US/10104271
; Publication NO. US20030181658A1
; GENERAL INFORMATION:
; APPLICANT: Madison, Edwin
; APPLICANT: Jiunn-Chern Yeh
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING SERINE PROTEASE CVSP14, THE ENCOD
; FILE REFERENCE: 24745-1614
; CURRENT APPLICATION NUMBER: US/10/104,271
```

```
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: 60/278,166
; FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-104-271-2
```

```
Query Match 100.0%; Score 4681; DB 14; Length 855;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 855; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGS DRARKGGGPKDFGAGLKYNRHEKVNGLGEEVEFLPVNNVKVKEKHGGRWVLA 60
|
|
|
Db 1 MGS DRARKGGGPKDFGAGLKYNRHEKVNGLGEEVEFLPVNNVKVKEKHGGRWVLA 60
|
|
|
QY 61 VLIGLLVLLGIGFLVHLYQYRDVRVQKVFNGYMRITNENFVDAYENSNSTEFVSLASKV 120
|
|
|
Db 61 VLIGLLVLLGIGFLVHLYQYRDVRVQKVFNGYMRITNENFVDAYENSNSTEFVSLASKV 120
|
|
|
QY 121 KDALKLYSGVPFLGPHYKESAVTAFSEGSAVIAYWSEFSIPOHLVEEAERVAERVM 180
|
|
|
Db 121 KDALKLYSGVPFLGPHYKESAVTAFSEGSAVIAYWSEFSIPOHLVEEAERVAERVM 180
|
|
|
QY 181 LPPRARSLSKSFVVTSVVAFPTDSKTQRTQDNCSFGLHARGVELMRTTTPGFPDPSYP 240
|
|
|
Db 181 LPPRARSLSKSFVVTSVVAFPTDSKTQRTQDNCSFGLHARGVELMRTTTPGFPDPSYP 240
|
|
|
QY 241 HARCQWALRGDADSVLSLTFRSFDLASCDERSGDLVTYNTLSPMEPHALVOLCGTYP 300
|
|
|
Db 241 HARCQWALRGDADSVLSLTFRSFDLASCDERSGDLVTYNTLSPMEPHALVOLCGTYP 300
|
|
|
QY 301 YNLTFHSSQNVLLITLITNTERRHGPGFATFPQLPRMSSCGGLRKAQGTNSPYPGHY 360
|
|
|
Db 301 YNLTFHSSQNVLLITLITNTERRHGPGFATFPQLPRMSSCGGLRKAQGTNSPYPGHY 360
|
|
|
QY 361 PPNIDCTWNIIEVPNNQHVSKFVYLLPEPGVPAGTCKPKDYVEINGEKYCGERSQFV 420
|
|
|
Db 361 PPNIDCTWNIIEVPNNQHVSKFVYLLPEPGVPAGTCKPKDYVEINGEKYCGERSQFV 420
|
|
|
QY 421 NSNKITVRPHSDQSYTDTGFLAEYLSYDSSDPCPGQFTCTGRCIRKELRCDGWADCT 480
|
|
|
Db 421 NSNKITVRPHSDQSYTDTGFLAEYLSYDSSDPCPGQFTCTGRCIRKELRCDGWADCT 480
|
|
|
QY 481 SDELNCSCDAGHQFTCKNFKCKPLFWVCDSDVNDGDNDEQSCCPAQTFRCNSGKCLSK 540
|
|
|
Db 481 SDELNCSCDAGHQFTCKNFKCKPLFWVCDSDVNDGDNDEQSCCPAQTFRCNSGKCLSK 540
|
|
|
QY 541 SQQNGKDDCGDSDGDEASCPKVNVTCTKHTYRCLNGLCLSKGNPECDGKEDCDSDG 600
|
|
|
Db 541 SQQNGKDDCGDSDGDEASCPKVNVTCTKHTYRCLNGLCLSKGNPECDGKEDCDSDG 600
|
|
|
QY 601 DCDGLRSFTRQARVVGTDAGEWPMQVSLHALGQGHICGASLISPNWLVAACHYID 660
|
|
|
Db 601 DCDGLRSFTRQARVVGTDAGEWPMQVSLHALGQGHICGASLISPNWLVAACHYID 660
|
|
|
QY 661 DRGFRYSDPTQWTAFLGLHDQSORSAQGVQERRLRKRIISHPPFNDFTDYDIALLE 720
|
|
|
Db 661 DRGFRYSDPTQWTAFLGLHDQSORSAQGVQERRLRKRIISHPPFNDFTDYDIALLE 720
|
|
|
QY 721 AEYSWMVRPCLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTCENLL 780
|
|
|
Db 721 AEYSWMVRPCLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTCENLL 780
|
|
|
QY 781 PQQITPRMCMVGFSLGGVSDSCQDGGPLSSVEADGRIFQAGVSWGDCGAQRNKP 840
|
|
|
Db 781 PQQITPRMCMVGFSLGGVSDSCQDGGPLSSVEADGRIFQAGVSWGDCGAQRNKP 840
|
|
|
QY 841 RLPLFRDWIKENTGV 855
|
|
|
```

Db 841 RLPLFRDWIKENTGV 855

RESULT 8

US-10-147-211A-2

; Sequence 2, Application US/10147211A

; Publication No. US20030285900A1

; GENERAL INFORMATION:

; APPLICANT: Madison, Edward

; APPLICANT: Yeh, Jium-Chern

; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING A TRANSMEMBRANE SERINE PROTEASE 1

; TITLE OF INVENTION: ENCODED POLYPEPTIDES AND METHODS BASED THEREON

; FILE REFERENCE: 24745-1616

; CURRENT APPLICATION NUMBER: US/10/147,211A

; CURRENT FILING DATE: 2002-05-14

; PRIOR FILING DATE: 60/291,001

; NUMBER OF SEQ ID NOS: 23

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 2

; LENGTH: 855

; TYPE: PRT

; ORGANISM: Homo Sapien

US-10-147-211A-2

Query Match 100.0%; Score 4681; DB 15; Length 855;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 855; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGS DRARKGGGPKDFGAGLKYNSRHEKVNGLGEEGVFLPVNNVKKVEKHGGRWVLA 60

Db 1 MGS DRARKGGGPKDFGAGLKYNSRHEKVNGLGEEGVFLPVNNVKKVEKHGGRWVLA 60

QY 61 VLI GLLLVLLIGIFLVHQLQYRDVVRQKVFNGVMRITNENFVDAYNSNSTEFVSLASKV 120

Db 61 VLI GLLLVLLIGIFLVHQLQYRDVVRQKVFNGVMRITNENFVDAYNSNSTEFVSLASKV 120

QY 121 KOALKLILYSGVPLGYPHKESAVTAFSEGSVIAYYNSFISIPQHLVEAEARVMAEERVVM 180

Db 121 KOALKLILYSGVPLGYPHKESAVTAFSEGSVIAYYNSFISIPQHLVEAEARVMAEERVVM 180

QY 181 LPPRARSLSKSFVTSVVAFTDSTKTQRTQDNSCSFGLHARGVELMRFTTFCGPDSPYPA 240

Db 181 LPPRARSLSKSFVTSVVAFTDSTKTQRTQDNSCSFGLHARGVELMRFTTFCGPDSPYPA 240

QY 241 HARCOMALRGDADSVLSLTFRSFDLASCDERGSIDLVTYNTLSPMEPHALVOLCGTYPPS 300

Db 241 HARCOMALRGDADSVLSLTFRSFDLASCDERGSIDLVTYNTLSPMEPHALVOLCGTYPPS 300

QY 301 YNLTFHSSQNVLLITLITNTERRHPGFEATFFQLPRMSSCGRLRKAQGTFSNPPYPGHY 360

Db 301 YNLTFHSSQNVLLITLITNTERRHPGFEATFFQLPRMSSCGRLRKAQGTFSNPPYPGHY 360

QY 361 PPNIDCTWNI EVNPNHVKVSKFFYLLLEPGVPAGTCPKDYVEINGEKYCGERSQFVVT 420

Db 361 PPNIDCTWNI EVNPNHVKVSKFFYLLLEPGVPAGTCPKDYVEINGEKYCGERSQFVVT 420

QY 421 NSNKITVRPHSDQSYTDTGFLAEYLSYSDSDPCPGQFTCTGRCIRKELRCDGWADCTDH 480

Db 421 NSNKITVRPHSDQSYTDTGFLAEYLSYSDSDPCPGQFTCTGRCIRKELRCDGWADCTDH 480

QY 481 SDPLNCSCDAGHQTCKNFKCPKPLFWVCDVNDGDNDEQSCCPAQTFRCNSGKCLSK 540

Db 481 SDPLNCSCDAGHQTCKNFKCPKPLFWVCDVNDGDNDEQSCCPAQTFRCNSGKCLSK 540

QY 541 SQOCHGKDDCGSDSEASCPKNNVTCTKHTYRCLNGLCLSKGNPECDGKEDCDGSDSEK 600

Db 541 SQOCHGKDDCGSDSEASCPKNNVTCTKHTYRCLNGLCLSKGNPECDGKEDCDGSDSEK 600

QY 601 DCDCLGRSFTQARVVGVDADGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYID 660

Db 601 DCDCLGRSFTQARVVGVDADGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYID 660

QY 661 DRGFRYSPTQWTAFLGLHDQSRGAPGVQERRLKRIISHPFNFDFDYDIALLELEKP 720

Db 661 DRGFRYSPTQWTAFLGLHDQSRGAPGVQERRLKRIISHPFNFDFDYDIALLELEKP 720

QY 721 ABYSWVRPICLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKEIRVINQTTCE 780

Db 721 ABYSWVRPICLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKEIRVINQTTCE 780

QY 781 POQITPRMVCVGLSGGVDSCQDGGPLSSVEADGRIFOAGVSWGDCGCAORNKPGVVT 840

Db 781 POQITPRMVCVGLSGGVDSCQDGGPLSSVEADGRIFOAGVSWGDCGCAORNKPGVVT 840

QY 841 RLPLFRDWIKENTGV 855

Db 841 RLPLFRDWIKENTGV 855

RESULT 9

US-10-156-214A-2

; Sequence 2, Application US/10156214A

; Publication No. US20040001801A1

; GENERAL INFORMATION:

; APPLICANT: Edwin L. Madison

; APPLICANT: Joseph Edward Semple

; APPLICANT: George P. Vlaeuk

; APPLICANT: Scott Jeffrey Kemp

; APPLICANT: Mallareddy Komandla

; APPLICANT: Daniel Vanna Siev

; TITLE OF INVENTION: Conjugates Activated By Cell Surface Proteases and Therapeutic

; TITLE OF INVENTION: Thereof

; FILE REFERENCE: 24745-1611

; CURRENT APPLICATION NUMBER: US/10/156,214A

; CURRENT FILING DATE: 2002-05-23

; NUMBER OF SEQ ID NOS: 611

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 2

; LENGTH: 855

; TYPE: PRT

; ORGANISM: Homo Sapien

US-10-156-214A-2

Query Match 100.0%; Score 4681; DB 15; Length 855;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 855; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGS DRARKGGGPKDFGAGLKYNSRHEKVNGLGEEGVFLPVNNVKKVEKHGGRWVLA 60

Db 1 MGS DRARKGGGPKDFGAGLKYNSRHEKVNGLGEEGVFLPVNNVKKVEKHGGRWVLA 60

QY 61 VLI GLLLVLLIGIFLVHQLQYRDVVRQKVFNGVMRITNENFVDAYNSNSTEFVSLASKV 120

Db 61 VLI GLLLVLLIGIFLVHQLQYRDVVRQKVFNGVMRITNENFVDAYNSNSTEFVSLASKV 120

QY 121 KOALKLILYSGVPLGYPHKESAVTAFSEGSVIAYYNSFISIPQHLVEAEARVMAEERVVM 180

Db 121 KOALKLILYSGVPLGYPHKESAVTAFSEGSVIAYYNSFISIPQHLVEAEARVMAEERVVM 180

QY 181 LPPRARSLSKSFVTSVVAFTDSTKTQRTQDNSCSFGLHARGVELMRFTTFCGPDSPYPA 240

Db 181 LPPRARSLSKSFVTSVVAFTDSTKTQRTQDNSCSFGLHARGVELMRFTTFCGPDSPYPA 240

QY 241 HARCOMALRGDADSVLSLTFRSFDLASCDERGSIDLVTYNTLSPMEPHALVOLCGTYPPS 300

Db 241 HARCOMALRGDADSVLSLTFRSFDLASCDERGSIDLVTYNTLSPMEPHALVOLCGTYPPS 300

QY 301 YNLTFHSSQNVLLITLITNTERRHPGFEATFFQLPRMSSCGRLRKAQGTFSNPPYPGHY 360

Db 301 YNLTFHSSQNVLLITLITNTERRHPGFEATFFQLPRMSSCGRLRKAQGTFSNPPYPGHY 360

QY 361 PPNIDCTWNI EVNPNHVKVSKFFYLLLEPGVPAGTCPKDYVEINGEKYCGERSQFVVT 420

Db 361 PPNIDCTWNI EVNPNHVKVSKFFYLLLEPGVPAGTCPKDYVEINGEKYCGERSQFVVT 420

QY 421 NSNKITVRFHSDQSYTDTGFLAEYLSYDSSDPCPGQFTCTGRCIRKELRCGWDGADCTDH 480
Db 421 NSNKITVRFHSDQSYTDTGFLAEYLSYDSSDPCPGQFTCTGRCIRKELRCGWDGADCTDH 480
QY 481 SDELNCSADAGHQTCKNFKCKPLFWVCDVNDGSDNSDEQSCSPAOTFRCSNGKCLSK 540
Db 481 SDELNCSADAGHQTCKNFKCKPLFWVCDVNDGSDNSDEQSCSPAOTFRCSNGKCLSK 540
QY 541 SQQCNKGKDDCGSDGDEASCPKVVNVCTKHTYRCLNGLCLSKGNPCDCKEDCSDGSDSK 600
Db 541 SQQCNKGKDDCGSDGDEASCPKVVNVCTKHTYRCLNGLCLSKGNPCDCKEDCSDGSDSK 600
QY 601 DCDGLRSFTRQARVVGTDADGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
Db 601 DCDGLRSFTRQARVVGTDADGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
QY 661 DRGFRYSDPTQWTAFLGLHDQORSAPGVQERLKEII SHPPFNDFTFYDIALLELEKP 720
Db 661 DRGFRYSDPTQWTAFLGLHDQORSAPGVQERLKEII SHPPFNDFTFYDIALLELEKP 720
QY 721 AEYSSMVRPCLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKEIRVINQTTCCENLL 780
Db 721 AEYSSMVRPCLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKEIRVINQTTCCENLL 780
QY 781 PQOITPRMCMVGLSGGVDSQCGDGGPLSSVEADGRIFQAGVWSWGDGCAQRNKGVT 840
Db 781 PQOITPRMCMVGLSGGVDSQCGDGGPLSSVEADGRIFQAGVWSWGDGCAQRNKGVT 840
QY 841 RLPLFRDWTIKENTGV 855
Db 841 RLPLFRDWTIKENTGV 855

RESULT 10

US-10-072-012-352

; Sequence 352, Application US/10072012

; Publication No. US2004003493A1

; GENERAL INFORMATION:

; APPLICANT: Tchernev, Velizar

; APPLICANT: Spytek, Kimberly

; APPLICANT: Zernhusen, Bryan

; APPLICANT: Patturajan, Meera

; APPLICANT: Rastelli, Luca

; APPLICANT: Shimkets, Richard

; APPLICANT: Li, Li

; APPLICANT: Gangolli, Baha

; APPLICANT: Padigar, Muralidhara

; APPLICANT: Anderson, David W.

; APPLICANT: Rastelli, Luca

; APPLICANT: Miller, Charles E.

; APPLICANT: Gerlach, Valerie

; APPLICANT: Taupier, Jr. Raymond J.

; APPLICANT: Gusev, Vladimir Y.

; APPLICANT: Colman, Steven D.

; APPLICANT: Wolenc, Adam R.

; APPLICANT: Pena, Carol E. A.

; APPLICANT: Furtak, Katarzyna

; APPLICANT: Grosse, William M.

; APPLICANT: Alsobrook II, John P.

; APPLICANT: Lepley, Denise M.

; APPLICANT: Rieger, Daniel K.

; APPLICANT: Burgess, Catherine E.

; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same

; FILE REFERENCE: 21402-258

; CURRENT APPLICATION NUMBER: US/10/072,012

; CURRENT FILING DATE: 2002-01-31

; PRIOR APPLICATION NUMBER: 60/265,102

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: 60/265,514

; PRIOR FILING DATE: 2001-01-31

; PRIOR APPLICATION NUMBER: 60/265,517

; PRIOR FILING DATE: 2001-01-31

; PRIOR APPLICATION NUMBER: 60/265,412

; PRIOR FILING DATE: 2001-01-31

; PRIOR APPLICATION NUMBER: 60/265,395
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/266,406
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/266,767
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/267,057
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/266,975
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/267,459
; PRIOR FILING DATE: 2001-02-08
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1391
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 352
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-072-012-352

Query Match 100.0%; Score 4681; DB 15; Length 855;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 855; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGSDRARKGGGPKDQFGAGLKYNSRHEKVNGLKEGVFLPVNNVKKVKGHGRWVWVLA 60

Db 1 MGSDRARKGGGPKDQFGAGLKYNSRHEKVNGLKEGVFLPVNNVKKVKGHGRWVWVLA 60

QY 61 VLIGLLLVLLGIGFLVHLOYRDVRVQKVFNGYMRITNENFVDAYENSNSTEFVSLASKV 120

Db 61 VLIGLLLVLLGIGFLVHLOYRDVRVQKVFNGYMRITNENFVDAYENSNSTEFVSLASKV 120

QY 121 KDALKLLYSYVPFLGPHYKESAVTAFSEGSVIAYTWSEFSIPOHLVEAEARVMAEERVVM 180

Db 121 KDALKLLYSYVPFLGPHYKESAVTAFSEGSVIAYTWSEFSIPOHLVEAEARVMAEERVVM 180

QY 181 LPPRARSLSFVWTSVVAFPDTSKTQRTQNSCSFGLHARGVELMRFTTTCGFPDPSYPA 240

Db 181 LPPRARSLSFVWTSVVAFPDTSKTQRTQNSCSFGLHARGVELMRFTTTCGFPDPSYPA 240

QY 241 HARCQWALRGDADSVLSLTFRSFDLASCDERGSDLVTVYNTLSMPHEHALVOLCGTTPS 300

Db 241 HARCQWALRGDADSVLSLTFRSFDLASCDERGSDLVTVYNTLSMPHEHALVOLCGTTPS 300

QY 301 YNLTFHSSQNVLLITLINTERRHPGFEATFQLPMSSCGRLRKAQGTNSPYPGHY 360

Db 301 YNLTFHSSQNVLLITLINTERRHPGFEATFQLPMSSCGRLRKAQGTNSPYPGHY 360

QY 361 PPNIDCTWNIIEVPNNQHVVSFKFYLLEPGVPAGTCKDYVEINGEKYCGERSQFVWTS 420

Db 361 PPNIDCTWNIIEVPNNQHVVSFKFYLLEPGVPAGTCKDYVEINGEKYCGERSQFVWTS 420

QY 421 NSNKITVRFHSDQSYTDTGFLAEYLSYDSSDPCPGQFTCTGRCIRKELRCGWDGADCTDH 480

Db 421 NSNKITVRFHSDQSYTDTGFLAEYLSYDSSDPCPGQFTCTGRCIRKELRCGWDGADCTDH 480

QY 481 SDELNCSADAGHQTCKNFKCKPLFWVCDVNDGSDNSDEQSCSPAOTFRCSNGKCLSK 540

Db 481 SDELNCSADAGHQTCKNFKCKPLFWVCDVNDGSDNSDEQSCSPAOTFRCSNGKCLSK 540

QY 541 SQQCNKGKDDCGSDGDEASCPKVVNVCTKHTYRCLNGLCLSKGNPCDCKEDCSDGSDSK 600

Db 541 SQQCNKGKDDCGSDGDEASCPKVVNVCTKHTYRCLNGLCLSKGNPCDCKEDCSDGSDSK 600

QY 601 DCDGLRSFTRQARVVGTDADGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYID 660

Db 601 DCDGLRSFTRQARVVGTDADGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYID 660

QY 661 DRGFRYSDPTQWTAFLGLHDQORSAPGVQERLKEII SHPPFNDFTFYDIALLELEKP 720

Db 661 DRGFRYSDPTQWTAFLGLHDQORSAPGVQERLKEII SHPPFNDFTFYDIALLELEKP 720

QY 721 AYSVMVRPCLPDPASHVFPAGKAIWVTGHTQYGGTALILQKGEIRVINQTTCCENLL 780
 Db 721 AYSVMVRPCLPDPASHVFPAGKAIWVTGHTQYGGTALILQKGEIRVINQTTCCENLL 780
 QY 781 PQQITPRMCMVGLSGVDSQCGSDGSLSSVEADGRIFQAGVVSWGDCQAQRNKPQYVT 840
 Db 781 PQQITPRMCMVGLSGVDSQCGSDGSLSSVEADGRIFQAGVVSWGDCQAQRNKPQYVT 840
 QY 841 RLPLFRDWIKENTGV 855
 Db 841 RLPLFRDWIKENTGV 855

RESULT 11

US-10-072-012-411
 ; Sequence 411, Application US/10072012
 ; Publication No. US20040053493A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Tchernev, Velizar
 ; APPLICANT: Spytek, Kimberly
 ; APPLICANT: Zerhusen, Bryan
 ; APPLICANT: Patturajan, Meera
 ; APPLICANT: Shimkets, Richard
 ; APPLICANT: Li, Li
 ; APPLICANT: Gangolli, Bha
 ; APPLICANT: Padigaru, Muralidhara
 ; APPLICANT: Anderson, David W.
 ; APPLICANT: Rastelli, Luca
 ; APPLICANT: Miller, Charles E.
 ; APPLICANT: Gerlach, Valerie
 ; APPLICANT: Taupier Jr, Raymond J.
 ; APPLICANT: Gusev, Vladimir Y.
 ; APPLICANT: Colman, Steven D.
 ; APPLICANT: Wolenc, Adam R.
 ; APPLICANT: Pena, Carol E. A
 ; APPLICANT: Furtak, Katarzyna
 ; APPLICANT: Grosse, William M.
 ; APPLICANT: Alsobrook II, John P.
 ; APPLICANT: Lepley, Denise M.
 ; APPLICANT: Rieger, Daniel K.
 ; APPLICANT: Burgess, Catherine E.
 ; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
 ; FILE REFERENCE: 21402-258
 ; CURRENT APPLICATION NUMBER: US/10/072,012
 ; CURRENT FILING DATE: 2002-01-31
 ; PRIOR APPLICATION NUMBER: 60/265,102
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: 60/265,514
 ; PRIOR FILING DATE: 2001-01-31
 ; PRIOR APPLICATION NUMBER: 60/265,517
 ; PRIOR FILING DATE: 2001-01-31
 ; PRIOR APPLICATION NUMBER: 60/265,412
 ; PRIOR FILING DATE: 2001-01-31
 ; PRIOR APPLICATION NUMBER: 60/265,395
 ; PRIOR FILING DATE: 2001-01-31
 ; PRIOR APPLICATION NUMBER: 60/266,406
 ; PRIOR FILING DATE: 2001-02-02
 ; PRIOR APPLICATION NUMBER: 60/266,767
 ; PRIOR FILING DATE: 2001-02-05
 ; PRIOR APPLICATION NUMBER: 60/267,057
 ; PRIOR FILING DATE: 2001-02-07
 ; PRIOR APPLICATION NUMBER: 60/266,975
 ; PRIOR FILING DATE: 2001-02-07
 ; PRIOR APPLICATION NUMBER: 60/267,459
 ; PRIOR FILING DATE: 2001-02-08
 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 1391
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 411
 ; LENGTH: 855
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-072-012-411

Query Match 100.0%; Score 4681; DB 15; Length 855;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 855; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MGSDRARKGSGGPKDFGAGLKYNSRHEKVGLEBGFPLVNNVKKVKGGRWVWVLA 60
 Db 1 MGSDRARKGSGGPKDFGAGLKYNSRHEKVGLEBGFPLVNNVKKVKGGRWVWVLA 60
 QY 61 VLIIGLLVLLGIGLFWHLOYRDVRQVFNQYMRITNENFVDAYENSNSSTFVSLASKV 120
 Db 61 VLIIGLLVLLGIGLFWHLOYRDVRQVFNQYMRITNENFVDAYENSNSSTFVSLASKV 120
 QY 121 KDALKLYSGVPFLGPHYKESAVTAFSGSVIAYYWSFSPQHLVEAEARVMAEERVVM 180
 Db 121 KDALKLYSGVPFLGPHYKESAVTAFSGSVIAYYWSFSPQHLVEAEARVMAEERVVM 180
 QY 181 LPPRARSLSFVVTSVVAFPTDSKTQVTONSCSFGHLHARGVELMRFTTGGFPDPSYPA 240
 Db 181 LPPRARSLSFVVTSVVAFPTDSKTQVTONSCSFGHLHARGVELMRFTTGGFPDPSYPA 240
 QY 241 HARGQWALRGDADSVLSLTFERSFDLASCDERGSDDLVTYNTLSPMEPHALVOLCGTYPPS 300
 Db 241 HARGQWALRGDADSVLSLTFERSFDLASCDERGSDDLVTYNTLSPMEPHALVOLCGTYPPS 300
 QY 301 YNLTFHSSQNVLLITLITNERRHPGPEATFQOLPRMSSCGGRLRKAQGTNPSYPYGHY 360
 Db 301 YNLTFHSSQNVLLITLITNERRHPGPEATFQOLPRMSSCGGRLRKAQGTNPSYPYGHY 360
 QY 361 PPNIDCTWNIIEVPNNQHVKSFKFYLLLEPGVPAGTCKPKDYVEINGEKYGGERSQFVVT 420
 Db 361 PPNIDCTWNIIEVPNNQHVKSFKFYLLLEPGVPAGTCKPKDYVEINGEKYGGERSQFVVT 420
 QY 421 NSNKITVRFHSDQSYTDTGLAEVLSYSDSDPCQOFTCRGTCRIRKELRCOGWADCTDH 480
 Db 421 NSNKITVRFHSDQSYTDTGLAEVLSYSDSDPCQOFTCRGTCRIRKELRCOGWADCTDH 480
 QY 481 SDELNCSADAGHQFTCKNFKCKPLFWVCDSDVNDGDNDSDEQSCPAQTRFCSNGKCLSK 540
 Db 481 SDELNCSADAGHQFTCKNFKCKPLFWVCDSDVNDGDNDSDEQSCPAQTRFCSNGKCLSK 540
 QY 541 SQQCNKGKDCGSDGDEASCPKVVNTCTKTYRCINGLCLSGNPECDGKEDCSGSDSK 600
 Db 541 SQQCNKGKDCGSDGDEASCPKVVNTCTKTYRCINGLCLSGNPECDGKEDCSGSDSK 600
 QY 601 DDCGLRSFTRQARVVGTDADGEWPNQVSLHALGQGHICGASLISPNLWLSAAHCYID 660
 Db 601 DDCGLRSFTRQARVVGTDADGEWPNQVSLHALGQGHICGASLISPNLWLSAAHCYID 660
 QY 661 DRGFRYSDPTQWTAFLGLHDQSQRSAPGVQERRLKRITISHPFFNDFTFDYDIALLELEKP 720
 Db 661 DRGFRYSDPTQWTAFLGLHDQSQRSAPGVQERRLKRITISHPFFNDFTFDYDIALLELEKP 720
 QY 721 AYSVMVRPCLPDPASHVFPAGKAIWVTGHTQYGGTALILQKGEIRVINQTTCCENLL 780
 Db 721 AYSVMVRPCLPDPASHVFPAGKAIWVTGHTQYGGTALILQKGEIRVINQTTCCENLL 780
 QY 781 PQQITPRMCMVGLSGVDSQCGSDGSLSSVEADGRIFQAGVVSWGDCQAQRNKPQYVT 840
 Db 781 PQQITPRMCMVGLSGVDSQCGSDGSLSSVEADGRIFQAGVVSWGDCQAQRNKPQYVT 840
 QY 841 RLPLFRDWIKENTGV 855
 Db 841 RLPLFRDWIKENTGV 855

RESULT 12
 US-10-072-012-418
 ; Sequence 418, Application US/10072012
 ; Publication No. US2004003493A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Tchernev, Velizar
 ; APPLICANT: Spytek, Kimberly

```

; APPLICANT: Zerhusen, Bryan
; APPLICANT: Patturajan, Meera
; APPLICANT: Shimkets, Richard
; APPLICANT: Li, Li
; APPLICANT: Gangolli, Raha
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Anderson, David W.
; APPLICANT: Rastelli, Luca
; APPLICANT: Miller, Charles E.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Taupier, Jr. Raymond J.
; APPLICANT: Gusev, Vladimir Y.
; APPLICANT: Colman, Steven D.
; APPLICANT: Wolenc, Adam R.
; APPLICANT: Pena, Carol E. A.
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Grose, William M.
; APPLICANT: Alsobrook II, John P.
; APPLICANT: Lepley, Denise M.
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Burgess, Catherine E.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-258
; CURRENT APPLICATION NUMBER: US/10/072,012
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 60/265,102
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 60/265,514
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,412
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,395
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/266,406
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/266,767
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/267,057
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/266,975
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/267,459
; PRIOR FILING DATE: 2001-02-08
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1391
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 418
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-072-012-418

Query Match 100.0%; Score 4681; DB 15; Length 855;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 855; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGSRRKGGGKDPFGAGLKYNSHEKVGVEFLPNNVKKVEKHGPGRWVLA 60
Db 1 MGSRRKGGGKDPFGAGLKYNSHEKVGVEFLPNNVKKVEKHGPGRWVLA 60

Qy 61 VLIGLLVLLGTGFLVHLQVRDVRQVFNQYMRITNFNFDAVENSNSFEVSLASKV 120
Db 61 VLIGLLVLLGTGFLVHLQVRDVRQVFNQYMRITNFNFDAVENSNSFEVSLASKV 120

Qy 121 KDALKLLYSVPRLGPHYKESAVTAFSGSVIAYWSEFSIPQHLVEEAERVAEERVVM 180
Db 121 KDALKLLYSVPRLGPHYKESAVTAFSGSVIAYWSEFSIPQHLVEEAERVAEERVVM 180

Qy 181 LPPRARSLSFVTSVVAFPDTSKTQVTONQSCSFLHARGVELMRPTTGPFPDPSYP 240
Db 181 LPPRARSLSFVTSVVAFPDTSKTQVTONQSCSFLHARGVELMRPTTGPFPDPSYP 240

```

```

241 HARCOWALRGDADSVLSLTERSFDLASCDERGSDLVTVTNTLSPMEPHALVOLCGTYPBS 300
Db HARCOWALRGDADSVLSLTERSFDLASCDERGSDLVTVTNTLSPMEPHALVOLCGTYPBS 300

Qy 301 YNLTFHSSQNVLLITLITNTERRHPGFATFFQLPRMSSCGGRLRKAQOTFNFPYPGHY 360
Db 301 YNLTFHSSQNVLLITLITNTERRHPGFATFFQLPRMSSCGGRLRKAQOTFNFPYPGHY 360

Qy 361 PPNDICTWNIIEVPNNQHVKSFKPFYLLPDPGVPAGTCDPKDYVEINKEKYGGRSQFVVT 420
Db 361 PPNDICTWNIIEVPNNQHVKSFKPFYLLPDPGVPAGTCDPKDYVEINKEKYGGRSQFVVT 420

Qy 421 NSNKITVRFHSDQSYDTGFLAEVLSYDSSDPCQOFTCRGTGRCIRKELRCDGWADCTDH 480
Db 421 NSNKITVRFHSDQSYDTGFLAEVLSYDSSDPCQOFTCRGTGRCIRKELRCDGWADCTDH 480

Qy 481 SDELNCSCDAGHQFTCKNFKCKPLFWVCDSDVNDGDNSEQGCSCPAQTFRCNSGKCLSK 540
Db 481 SDELNCSCDAGHQFTCKNFKCKPLFWVCDSDVNDGDNSEQGCSCPAQTFRCNSGKCLSK 540

Qy 541 SQQNGKDDCGSDSEASCPKVNVTCTKHTYRCNLGLCLSKGNPECDGKEDCSGSDSK 600
Db 541 SQQNGKDDCGSDSEASCPKVNVTCTKHTYRCNLGLCLSKGNPECDGKEDCSGSDSK 600

Qy 601 DCDGLRSFTQARVVGTDADGEWPMQVSLHALGQGHICGASLISPNMLVSAACVYD 660
Db 601 DCDGLRSFTQARVVGTDADGEWPMQVSLHALGQGHICGASLISPNMLVSAACVYD 660

Qy 661 DRGFRYSDPTQWTAFLGLHDQSORSAFGVQRRRLKRIISHPFFNDFTFDYDIALLELEKP 720
Db 661 DRGFRYSDPTQWTAFLGLHDQSORSAFGVQRRRLKRIISHPFFNDFTFDYDIALLELEKP 720

Qy 721 AEYSSMVRPCLPDASHVFPAGKAIWVTGMWGTQYGGTALILQKGEIRVINOTTENLL 780
Db 721 AEYSSMVRPCLPDASHVFPAGKAIWVTGMWGTQYGGTALILQKGEIRVINOTTENLL 780

Qy 781 PQQITPRMCMVGLSGGVDSQSGDSGGPLSSVEADGRIFQAGVYVSWGDCQARNKPGVYT 840
Db 781 PQQITPRMCMVGLSGGVDSQSGDSGGPLSSVEADGRIFQAGVYVSWGDCQARNKPGVYT 840

Qy 841 RLPLFRDWIKENTGV 855
Db 841 RLPLFRDWIKENTGV 855

```

```

RESULT 13
US-10-600-187-2
; Sequence 2, Application US/10600187
; Publication No. US20040086910A1
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Tanimoto, Hirotooshi
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease
; FILE REFERENCE: D6064CIP/D
; CURRENT APPLICATION NUMBER: US/10/600,187
; PRIOR APPLICATION NUMBER: US/09/654,600A
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 09/421,213
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 2
; TYPE: PRT
; LENGTH: 855
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: TADG-15
US-10-600-187-2

```

```
Query Match 100.0%; Score 4681; DB 15; Length 855;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 855; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGS DRARKGGGPKDFGAGLKYNRSRHEKVNGLKEGVFLPVNNVKYKKGPGRWVLA 60
Db 1 MGS DRARKGGGPKDFGAGLKYNRSRHEKVNGLKEGVFLPVNNVKYKKGPGRWVLA 60

Qy 61 VLIGLLLVLLGIGFLVHLOQYRDVRVQKVFNGYMRITNENFVDAYENSNSFEVSLASKV 120
Db 61 VLIGLLLVLLGIGFLVHLOQYRDVRVQKVFNGYMRITNENFVDAYENSNSFEVSLASKV 120

Qy 121 KDALKLLSYGVPLPGPYHKESAVTAFSEGSVIAYWSEFSIPQHLVEAEERVAERVM 180
Db 121 KDALKLLSYGVPLPGPYHKESAVTAFSEGSVIAYWSEFSIPQHLVEAEERVAERVM 180

Qy 181 LPPRARSLSKSFVVTSSVAFPTDSKTVQTDNSCSFGLHARGVELMRFTTFCGPDSPYPA 240
Db 181 LPPRARSLSKSFVVTSSVAFPTDSKTVQTDNSCSFGLHARGVELMRFTTFCGPDSPYPA 240

Qy 241 HARCQWALRGDADSVLSLTFRSFDLASCDERSGDLVTYNTLSMPHEHALVOLCGTYP 300
Db 241 HARCQWALRGDADSVLSLTFRSFDLASCDERSGDLVTYNTLSMPHEHALVOLCGTYP 300

Qy 301 YNLTFHSSQNVLLITLITNTERRHPGFEATFQQLPRMSSCGRLRKAQGTNSPYPGHY 360
Db 301 YNLTFHSSQNVLLITLITNTERRHPGFEATFQQLPRMSSCGRLRKAQGTNSPYPGHY 360

Qy 361 PNIDCTWNIEVPNNQHVSKPKFYLLEPGVPAGTCKPDYVEINGEKYCGERSQFVVT 420
Db 361 PNIDCTWNIEVPNNQHVSKPKFYLLEPGVPAGTCKPDYVEINGEKYCGERSQFVVT 420

Qy 421 NSNKITVRPHSDQSYTDTGFLAEYLSYDSSDPCPGQFTCRGRCIRKELRCDGWADCTDH 480
Db 421 NSNKITVRPHSDQSYTDTGFLAEYLSYDSSDPCPGQFTCRGRCIRKELRCDGWADCTDH 480

Qy 481 SDELNCSCDAGHQTCKNFKCKPLFWVCDSVNDCGNSDEQGCSCPAQTFRCNSGKCLSK 540
Db 481 SDELNCSCDAGHQTCKNFKCKPLFWVCDSVNDCGNSDEQGCSCPAQTFRCNSGKCLSK 540

Qy 541 SQQNGKDDCGDSDEASCPKVNVTCTKHTYRCLNGCLSKGNPECDGKEDCSGDSDEK 600
Db 541 SQQNGKDDCGDSDEASCPKVNVTCTKHTYRCLNGCLSKGNPECDGKEDCSGDSDEK 600

Qy 601 DCDCGLRSFTRQARVVGTDADGEWPMQVSLHALGQCHI CGASLISPNWLVSAAHCYID 660
Db 601 DCDCGLRSFTRQARVVGTDADGEWPMQVSLHALGQCHI CGASLISPNWLVSAAHCYID 660

Qy 661 DRGFYSDPTQWTAFLGLHDQSQRSAPGVQERRLKRIISHPFFNDFTFDYDIALLELEKP 720
Db 661 DRGFYSDPTQWTAFLGLHDQSQRSAPGVQERRLKRIISHPFFNDFTFDYDIALLELEKP 720

Qy 721 ABEYSSMVRPCLPDASHVFPAGKAIWVTGHWGHTQYGGTGALILQKGEIRVINQTTCE 780
Db 721 ABEYSSMVRPCLPDASHVFPAGKAIWVTGHWGHTQYGGTGALILQKGEIRVINQTTCE 780

Qy 781 PQQITPRMCMVGLSGGVSDSCQDSSGLSSVEADGRIFQAGVSWGSDGCAQRNKP 840
Db 781 PQQITPRMCMVGLSGGVSDSCQDSSGLSSVEADGRIFQAGVSWGSDGCAQRNKP 840

Qy 841 RLPLFRDWIKENTGV 855
Db 841 RLPLFRDWIKENTGV 855
```

RESULT 14

```
US-10-612-466B-2
; Sequence 2, Application US/10612466B
; Publication No. US20050112579A1
; GENERAL INFORMATION:
; APPLICANT: Madison, Edwin
; APPLICANT: Ong, Edgar
; APPLICANT: Yeh, Juinn-Chern
```

```
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING SERINE PROTEASE 16, THE
; TITLE OF INVENTION: ENCODED PROTEINS AND METHODS BASED THEREON
; FILE REFERENCE: 24745-1625
; CURRENT APPLICATION NUMBER: US/10/612,466B
; CURRENT FILING DATE: 2003-07-01
; PRIOR APPLICATION NUMBER: 60/394,347
; PRIOR FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-612-466B-2
```

```
Query Match 100.0%; Score 4681; DB 17; Length 855;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 855; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 MGS DRARKGGGPKDFGAGLKYNRSRHEKVNGLKEGVFLPVNNVKYKKGPGRWVLA 60
Db 1 MGS DRARKGGGPKDFGAGLKYNRSRHEKVNGLKEGVFLPVNNVKYKKGPGRWVLA 60

Qy 61 VLIGLLLVLLGIGFLVHLOQYRDVRVQKVFNGYMRITNENFVDAYENSNSFEVSLASKV 120
Db 61 VLIGLLLVLLGIGFLVHLOQYRDVRVQKVFNGYMRITNENFVDAYENSNSFEVSLASKV 120

Qy 121 KDALKLLSYGVPLPGPYHKESAVTAFSEGSVIAYWSEFSIPQHLVEAEERVAERVM 180
Db 121 KDALKLLSYGVPLPGPYHKESAVTAFSEGSVIAYWSEFSIPQHLVEAEERVAERVM 180

Qy 181 LPPRARSLSKSFVVTSSVAFPTDSKTVQTDNSCSFGLHARGVELMRFTTFCGPDSPYPA 240
Db 181 LPPRARSLSKSFVVTSSVAFPTDSKTVQTDNSCSFGLHARGVELMRFTTFCGPDSPYPA 240

Qy 241 HARCQWALRGDADSVLSLTFRSFDLASCDERSGDLVTYNTLSMPHEHALVOLCGTYP 300
Db 241 HARCQWALRGDADSVLSLTFRSFDLASCDERSGDLVTYNTLSMPHEHALVOLCGTYP 300

Qy 301 YNLTFHSSQNVLLITLITNTERRHPGFEATFQQLPRMSSCGRLRKAQGTNSPYPGHY 360
Db 301 YNLTFHSSQNVLLITLITNTERRHPGFEATFQQLPRMSSCGRLRKAQGTNSPYPGHY 360

Qy 361 PNIDCTWNIEVPNNQHVSKPKFYLLEPGVPAGTCKPDYVEINGEKYCGERSQFVVT 420
Db 361 PNIDCTWNIEVPNNQHVSKPKFYLLEPGVPAGTCKPDYVEINGEKYCGERSQFVVT 420

Qy 421 NSNKITVRPHSDQSYTDTGFLAEYLSYDSSDPCPGQFTCRGRCIRKELRCDGWADCTDH 480
Db 421 NSNKITVRPHSDQSYTDTGFLAEYLSYDSSDPCPGQFTCRGRCIRKELRCDGWADCTDH 480

Qy 481 SDELNCSCDAGHQTCKNFKCKPLFWVCDSVNDCGNSDEQGCSCPAQTFRCNSGKCLSK 540
Db 481 SDELNCSCDAGHQTCKNFKCKPLFWVCDSVNDCGNSDEQGCSCPAQTFRCNSGKCLSK 540

Qy 541 SQQNGKDDCGDSDEASCPKVNVTCTKHTYRCLNGCLSKGNPECDGKEDCSGDSDEK 600
Db 541 SQQNGKDDCGDSDEASCPKVNVTCTKHTYRCLNGCLSKGNPECDGKEDCSGDSDEK 600

Qy 601 DCDCGLRSFTRQARVVGTDADGEWPMQVSLHALGQCHI CGASLISPNWLVSAAHCYID 660
Db 601 DCDCGLRSFTRQARVVGTDADGEWPMQVSLHALGQCHI CGASLISPNWLVSAAHCYID 660

Qy 661 DRGFYSDPTQWTAFLGLHDQSQRSAPGVQERRLKRIISHPFFNDFTFDYDIALLELEKP 720
Db 661 DRGFYSDPTQWTAFLGLHDQSQRSAPGVQERRLKRIISHPFFNDFTFDYDIALLELEKP 720

Qy 721 ABEYSSMVRPCLPDASHVFPAGKAIWVTGHWGHTQYGGTGALILQKGEIRVINQTTCE 780
Db 721 ABEYSSMVRPCLPDASHVFPAGKAIWVTGHWGHTQYGGTGALILQKGEIRVINQTTCE 780

Qy 781 PQQITPRMCMVGLSGGVSDSCQDSSGLSSVEADGRIFQAGVSWGSDGCAQRNKP 840
Db 781 PQQITPRMCMVGLSGGVSDSCQDSSGLSSVEADGRIFQAGVSWGSDGCAQRNKP 840
```

Db 781 PQ0ITPRMCMVGLSGGVDSQCGSGPLSSVEADGRIFQAGVSWGDCGAQRNKPQVYT 840
Qy 841 RLPLFRDWIKENTGV 855
Db 841 RLPLFRDWIKENTGV 855
RESULT 15
US-10-295-027-1185
; Sequence 1185, Application US/10295027
; Publication No. US20030212350A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Nataasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynn, Richard
; APPLICANT: Hevezi, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; TITLE OF INVENTION: Methods of Screening for Modulators of Cancer
; FILE REFERENCE: 018501-012500US
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US/10/295,027
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1185
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-295-027-1185
Query Match 99.9%; Score 4676; DB 15; Length 855;
Best Local Similarity 99.9%; Pred No. 0;
Matches 854; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 MGSRRARKGGGPKDFGAGLKYNSRHEKVNGLGVEFLPVNNVKKVEKHGFGRWVLA 60
Db 1 MGSRRARKGGGPKDFGAGLKYNSRHEKVNGLGVEFLPVNNVKKVEKHGFGRWVLA 60
Qy 61 VLIGLLLVLLGIGFLVWLHQLRDVRVQVFNQYMRITNFENFDAYENSNSTEFVSLASKV 120
Db 61 VLIGLLLVLLGIGFLVWLHQLRDVRVQVFNQYMRITNFENFDAYENSNSTEFVSLASKV 120
Qy 121 KDALKLLYSGVPFLGPHYKESAVTAFSGSVIAYVWSFSPQHLYVEAEERVAEERVVM 180
Db 121 KDALKLLYSGVPFLGPHYKESAVTAFSGSVIAYVWSFSPQHLYVEAEERVAEERVVM 180
Qy 181 LPPRARSLSKSFVTSVVAFTDSKTQVQDQNSCSFGLHARGVELMRFTTGPFPDSPA 240

Db 181 LPPRARSLSKSFVTSVVAFTDSKTQVQDQNSCSFGLHARGVELMRFTTGPFPDSPA 240
Qy 241 HARCQWALRGDADSVLSLTFRSFDLASCDESGDLVTYNTLSNMEPHALVOLCGTYP 300
Db 241 HARCQWALRGDADSVLSLTFRSFDLASCDESGDLVTYNTLSNMEPHALVOLCGTYP 300
Qy 301 YNLTFHSSQNVLLITLINTERRHPGFATFQQLPRMSSCGRLRKAQGTFSNPPYGHY 360
Db 301 YNLTFHSSQNVLLITLINTERRHPGFATFQQLPRMSSCGRLRKAQGTFSNPPYGHY 360
Qy 361 PPNIDCTWNIIEVPNNQHVKSFKPFYLLPEGPVAGTCKPKYVEINGEKYGGERSQFV 420
Db 361 PPNIDCTWNIIEVPNNQHVKSFKPFYLLPEGPVAGTCKPKYVEINGEKYGGERSQFV 420
Qy 421 NSNKITVRFHSDQSYTDTGFLAEVLSYDSSDPCPGQFTCRGRCIRKELRCDGWADCD 480
Db 421 NSNKITVRFHSDQSYTDTGFLAEVLSYDSSDPCPGQFTCRGRCIRKELRCDGWADCD 480
Qy 481 SDELNCSCDAGHQFTCKNFKCKPLFWVCDVNDGDNDSDEQSCCPAOTFRCSNGKCLSK 540
Db 481 SDELNCSCDAGHQFTCKNFKCKPLFWVCDVNDGDNDSDEQSCCPAOTFRCSNGKCLSK 540
Qy 541 SQQNGKDDCGDSDASCPKVVVTCYKTYRCINGLCLSKNPECDCKEDCSDGSD 600
Db 541 SQQNGKDDCGDSDASCPKVVVTCYKTYRCINGLCLSKNPECDCKEDCSDGSD 600
Qy 601 DDCGLRSFTRQARVVGTDADGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
Db 601 DDCGLRSFTRQARVVGTDADGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
Qy 661 DRGFYSDPTQWTAFLGLHDQSORSAFGVQERRLRIISHPFFNDFTFDYDIALLELEKP 720
Db 661 DRGFYSDPTQWTAFLGLHDQSORSAFGVQERRLRIISHPFFNDFTFDYDIALLELEKP 720
Qy 721 AEYSSMVRPICLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTTCE 780
Db 721 AEYSSMVRPICLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTTCE 780
Qy 781 PQ0ITPRMCMVGLSGGVDSQCGSGPLSSVEADGRIFQAGVSWGDCGAQRNKPQVYT 840
Db 781 PQ0ITPRMCMVGLSGGVDSQCGSGPLSSVEADGRIFQAGVSWGDCGAQRNKPQVYT 840
Qy 841 RLPLFRDWIKENTGV 855
Db 841 RLPLFRDWIKENTGV 855

Search completed: September 23, 2005, 13:12:20
Job time : 178 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 23, 2005, 12:48:10 ; Search time 23 Seconds
(without alignments)
3576.754 Million cell updates/sec

Title: US-09-421-213-2
Perfect score: 4681
Sequence: 1 MGSRRKGGGPKDFGAGL.....PGVTRPLPRDWIKENTGV 855

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_79.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3883	83.0	855	2 JC7731	membrane-bound arg
2	727	15.5	1034	1 A53663	enteropeptidase (E
3	712.5	15.2	1035	1 A43090	enteropeptidase (E
4	682.5	14.6	1019	1 A56318	enteropeptidase (E
5	663.5	14.2	1113	2 JE0315	low-density lipopr
6	578.5	12.4	1524	2 T30337	polyprotein - Afri
7	574	12.3	699	1 I54763	Ra-reactive factor
8	533.5	11.4	638	1 KQMSPL	plasma kallikrein
9	514.5	11.0	638	1 KQRTPL	plasma kallikrein
10	509.5	10.9	790	1 PLPG	plasma kallikrein
11	506	10.8	613	2 S15468	complement C3b/C4b
12	502	10.7	460	2 B61545	plasma (EC 3.4.21
13	501.5	10.7	786	1 A47547	serine proteinase
14	500	10.7	638	1 KQHUP	plasma kallikrein
15	497	10.6	810	1 PLHU	plasma (EC 3.4.21
16	492.5	10.5	583	2 A29154	complement factor
17	491.5	10.5	812	1 PLMS	plasma (EC 3.4.21
18	491	10.5	416	1 KFB0	coagulation factor
19	490.5	10.5	812	1 PLB0	plasma (EC 3.4.21
20	490	10.5	417	1 S00845	hepsin (EC 3.4.21
21	486	10.4	810	2 B30848	plasma (EC 3.4.21
22	484.5	10.4	343	1 A57014	proctasin (EC 3.4
23	482	10.3	686	1 A59271	Ra-reactive factor
24	480.5	10.3	416	1 S33777	hepsin (EC 3.4.21
25	480	10.3	625	1 KFHU1	coagulation factor
26	479.5	10.2	461	1 KFHU	coagulation factor
27	478	10.2	810	2 I46260	plasma (EC 3.4.21
28	472.5	10.1	2616	2 A57096	nudel protein prec
29	472	10.1	275	2 S40007	trypsin (EC 3.4.21

RESULT 1

JC7731

membrane-bound arginine-specific serine proteinase precursor - rat

C;Species: Rattus norvegicus (Norway rat)

C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004

C;Accession: JC7731; JC7775

R;Kishi, K.; Yamazaki, K.; Yasuda, I.; Yahagi, N.; Ichinose, M.; Tsuchiya, Y.; Athauda

J. Biochem. 130, 425-430, 2001

A;Title: Characterization of a membrane-bound arginine-specific serine protease from r

A;Reference number: JC7731; MUID:21421307; PMID:11530019

A;Accession: JC7731

A;Molecule type: mRNA

A;Residues: 1-855 <KIS>

A;Cross-references: UNIPROT:Q9JUI7; DDBJ:AB049189

A;Experimental source: strain Male, 7-week-old

R;Satomi, S.; Yanasaki, Y.; Tezuka, S.; Hitomi, Y.; Iwanaga, T.; Fushiki, T.

Biochem. Biophys. Res. Commun. 287, 995-1002, 2001

A;Title: A role for membrane-type serine protease (MT-Sp1) in intestinal epithelial tu

A;Reference number: JC7775; PMID:11573963

A;Contents: Small intestine

A;Accession: JC7775

A;Molecule type: mRNA

A;Residues: 1-855 <SAT>

A;Cross-references: DDBJ:AB037898

C;Comment: This enzyme, an epithelial-derived, type II integral membrane serine protease of specific proteins or peptides on the brushborder membranes. It also participates i: lial migration and/or cell loss.

C;Genetics:

A;Gene: mt-sp1

A;Map position: basolateral cell surface

C;Superfamily: membrane-bound arginine-specific serine proteinase

C;Keywords: protein digestion

Query Match 83.0%; Score 3883; DB 2; Length 855;
Best Local Similarity 81.1%; Pred. No. 7.6e-246;
Matches 693; Conservative 79; Mismatches 83; Indels 0; Gaps 0;

Qy 1 MGSRRKGGGPKDFGAGLKYNRHKVNGLEGEVEFLPVNNVKVKEKPGRWVLA 60

Db 1 MGNRRKAGGGGQDFGAGLKYNRLENMNGFEVEFLPVNNAKQVEKGPWWVMA 60

Qy 61 VLIGLLLVLLGIGFLVHLYQYRDVVRQKVFNGVMYRTNENFDAYENSTFEFVLSKV 120

Db 61 VVPSFLLLSLMAGLLVHFRNVRQKVFNGHLRTNENFDAYENSTFEFVLSQV 120

Qy 121 KDALKLLYSGVPLPGPYHKESAVTAFSEGSVIAYTWSEFSIPQHLVBEARVMAEERVM 180

Db 121 KEALKLWYSEVPVLGPYHKSTVTAASEGSVIAYTWSEFSIPHLVEEDVMAVERVVT 180

Qy 181 LPPRSLKSFVVTSVVAFPTDSKTVQRTQDNCSFGLHARGVELMRFTTTPGPDSPYA 240

Db 181 LPPRRLALKSVLTSSVAFPTDPRMLQRTQDNCSFGLHARGVTRFTTTPGFPNSPYA 240

coagulation factor
plasmin (EC 3.4.21
trypsin (EC 3.4.21
t-plasminogen acti
trypsin (EC 3.4.21
trypsin (EC 3.4.21
apolipoprotein(a)
plasma hyaluronan-
trypsin-like prote
apoprotein(a) (EC
oviductin (EC 3.4.
trypsin-like prote
mast cell tryptase
trypsin (EC 3.4.21
trypsin (EC 3.4.21
chymotrypsin (EC 3

ALIGNMENTS

Db 794 QEVSKIVGNDSDREGAPWVALYNGQ-LLCGASLVSRDLVSAACHCVYG-----RNLE 848
Qy 669 PTQWTAFLGLHDQSORSAFPGVQERLKRILSHPPFNDFDYDIALLELEKPAEYSSMVR 728
Db 849 PSKWKALGLHMTSNLTSPQIVTRILDELVINPHYNRRKSDIAMHLEFKVNTDYIQ 908
Qy 729 FICLPDASHVFPAGKAIWVTGHTQYGGTGALILQKGEIRVINOTTENLLPQ-QITPR 787
Db 909 FICLPEENQVFPFGRICSTAGKVIYQSGPADILQEADVLLSNEKCOQOMPEYNITEN 968
Qy 788 MNCVGLSGVDSCGDSGGLSSVEADGRIFQAGVWSGDCQARNKGGVYTRLPFRD 847
Db 969 MNCAGYEGGIDSCGDSGGLMCLCENNRWLLAGTSGFYQCALPNRPGVYARVPKFE 1027
Qy 848 WIK 850
Db 1028 WIQ 1030
RESULT 3
A43090
enteropeptidase (EC 3.4.21.9) precursor [validated] - bovine
N:Alternate names: enterokinase
C:Species: Bos primigenius taurus (cattle)
C:Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text_change 09-Jul-2004
C:Accession: A43090; A48874; A61436
R:Kitamoto, Y.; Yuan, X.; Wu, Q.; McCourt, D.W.; Sadler, J.E.
Proc. Natl. Acad. Sci. U.S.A. 91, 7588-7592, 1994
A:Title: Enterokinase, the initiator of intestinal digestion, is a mosaic protease composed of two distinct regions
A:Reference number: A43090; MUID:94329561; PMID:8052624
A:Accession: A43090
A:Status: nucleic acid sequence not shown; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1035 <KIT>
A:Cross-references: UNIPROT:P98072; GB:U09859; NID:9746410; PIDN:AAB40026.1; PID:g746411
A:Experimental source: small intestine
R:LaVallie, E.R.; Rehemtulla, A.; Racie, L.A.; DiBlasio, E.A.; Ferenz, C.; Grant, K.L.; J. Biol. Chem. 268, 23311-23317, 1993
A:Title: Cloning and functional expression of a cDNA encoding the catalytic subunit of human enterokinase
A:Reference number: A48874; MUID:94043122; PMID:8226855
A:Accession: A48874
A:Molecule type: mRNA
A:Residues: 801-1035 <LAV>
A:Cross-references: GB:L19663; NID:9416131; PIDN:AAAL6035.1; PID:g416132
A:Note: parts of this sequence, including the amino end of the mature protein, were confirmed by sequencing of the complementary DNA
R:Light, A.; Janska, H. J. Protein Chem. 10, 475-480, 1991
A:Title: The amino-terminal sequence of the catalytic subunit of bovine enterokinase.
A:Reference number: A61436; MUID:92189715; PMID:1799406
A:Accession: A61436
A:Molecule type: protein
A:Residues: 801-807, Y', 809-827 <LIG>
C:Comment: The mechanism of association with the membrane of the intestinal brush border enzyme attachment using a signal-anchor sequence.
C:Comment: Conversion from membrane-bound to soluble forms may involve further processing
C:Complex: mature enteropeptidase is variously reported to contain two (heavy and light) l-fide linked
C:Function:
A:Description: cleaves propeptide from trypsinogen to produce active trypsin
A:Pathway: intestinal digestive hydrolase cascade
C:Superfamily: enteropeptidase; Clr/Cls repeat homology; LDL receptor ligand-binding repeat
C:Keywords: glycoprotein; hydrolase; intestine; serine proteinase; transmembrane protein
F:22-38/Domain: transmembrane #status predicted <TM>
F:52-117/Product: enteropeptidase mini chain #status predicted <MCH>
F:118-800/Product: enteropeptidase heavy chain #status predicted <HCH>
F:199-236/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F:358-520/Domain: MAM homology <MAM>
F:542-647/Domain: Clr/Cls repeat homology <CLR>
F:659-693/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F:694-799/Domain: scavenger receptor cysteine-rich domain homology #status atypical <SRC>
F:801-1035/Product: enteropeptidase light chain #status predicted <LCH>
F:801-1030/Domain: trypsin homology <TRY>

F:116,147,170,194,233,263,264,404,456,486,519,550,646,698,722,741,762,864,903,965/Binc
F:788-912,826-842,926-993,957-972,983-1011/Diculfide bonds: #status predicted
F:841,892,987/Active site: His, Asp, Ser #status predicted
Query Match 15.2%; Score 712.5; DB 1; Length 1035;
Best Local Similarity 23.3%; Pred. No. 1.2e-38;
Matches 253; Conservative 149; Mismatches 319; Indels 365; Gaps 45;
Qy 55 WVLAALVIGLILLLVLLGIGFLVWHLQVRDVRQKVF--NGYMR-ITNENRFDAYENST 111
Db 21 FAVLFLVILVALCGLIAVSWLSIQGSVKDAFAGSHKSGTGLKILSGATYVPHLOKLSV 80
Qy 112 EFVSLASKVDKALKLYSGVFPFLGPHYKHSAVTAFAFSEGVIAAY-----W-SFESIPQH 164
Db 81 DFKVLAFDIQMIDDIQSSNLKNEY-KNSRVLFQENGSIIVIFDLFLFDQWVSDQNVKEE 139
Qy 165 LVEEAERVMAEERVM-----LPPRSLKSFVTSV-----V 197
Db 140 LIOGIEANKSQLVTFHIDLNSIDITASLENFSTISATTSEK--LTTISPLATPGNYSI 197
Qy 198 APPTDSKTQVORTDQNSCSFGLHARGVELM-----RFTTPGF 233
Db 198 ECPDSSL-----CADALKCIAIDLFCDELNCPDGSDEDNKTATACDGRFLLTGS 249
Qy 234 PDS-----PYPAH---ARCOVALRGDADSVLSLTFRSFD----- 264
Db 250 SGSFEALHYPKPSNNTSAVCRWIIRVNOGLSIQLNFDYFNTYADVLNLYEGMGSSKILR 309
Qy 265 -----LASCDERGSDLV--TVNTLSMPHEHALVQL----- 293
Db 310 ASLWSNPNGIIRIFSNQVATFIQSDS--SDYIGFKVTVYAFNSKELNNYEKINCFED 367
Qy 294 --C-----GTYPPSYNLTF-HS----- 307
Db 368 GFCFWIQDLNDDNEWERTQSTFPSTGPTFDHTFGNESGFYSTPTGPGRRRVRGLLT 427
Qy 308 -----SONVLLITLTNTERRHPGFPEATPFQ----- 333
Db 428 LPLDPTPEQACLSFWYMYGCVNVKLSINISDQ---NMEKTIQKEGNYQNNWYQVGT 484
Qy 334 -----LPRMSSCGG 342
Db 485 LNETVEFKSVFYGKQNILSDIALDLSITGYICNVSVYPEPTLVPTPPPELP--TDCGG 542
Qy 343 --RLRKAQGFNSPYPGHYPPNIDCTWNTVEPNNOHVKSEKFFYLLERGVGPACTPKD 400
Db 543 PHDLWEPNTFTSINFPNSYNOAFCLWNLNAOKGKNIQLHFQFEDLENTA-----D 594
Qy 401 YVEI-NGEKYCGRRSQFV-----VTSNSNKITVRFHSDQSYTDITGFLAEYLS-- 446
Db 595 VVEIRDE---GDSLFLAVYTGPGVPNDVPSTNRTMTVLFTIDNMLAKQGFKANFTGY 651
Qy 447 -YDSSDPC-PGQFTCTGTCIRKELRCDCGWADCTDHSDELNC-----SCDAGH--OPTC 496
Db 652 GLGIPEPCEDNPKQDGEICPLVNLCDGPPHCKDGSDEAHCVRLFNFTDSSGLVQFRI 711
Qy 497 KNKFCPLFWCVSDVNDGNSDEQSCSPAQTPRCSNGKLSKSCQCNCKDCCGDSDE 556
Db 712 QS-----IWHV-----ACAEN-----WTTQISDDVC-----LLGNS 743
Qy 557 ASC-----PKNVVVTCTKHTYRCLNGLCLSKGNPECDGKDCGSDGDEKDCD---CGLR 607
Db 744 VPTFTSGGPPYVNLNAP-----NGSLILTPSQCC-----LEDLSILLQCNYSKCGK 791
Qy 608 SFTRQA--RVVGTGTDADGEPWQVSLHALGQGHICGASLISPNWLVSAAHCYIDDRGFR 665
Db 792 LVTOEVSPIKVGSDSREGAWPWWVALYFDQ-QVCCASLVSRDLVLSAAHCYVYG---R 846
Qy 666 YSDPTQWTAFLGLHDQSORSAFPGVQERLKRILSHPPFNDFDYDIALLELEKPAEYSS 725
Db 847 NMEPSKWKAVILGHMASNLTSPOIETRLDQIVINPHYNRRKNDIAMHLEMKVNYTD 906
Qy 726 MVRPICLPDASHVFPAGKAIWVTGHTQYGGTGALILQKGEIRVINOTTENLLPQ-QI 784

Db 907 YIPICLPENQYFPFCRICSIAGWALYQGSTADVLQEAADVPLLSNEKCCQOMPEYNI 966
QY 785 TPRLMVCVGLSGVDSQCGDGGPLSSVEADGRIFQAGVVSWGDCGAQNRKPGVYTRPLP 844
Db 967 TENMVCAGYENGVDSCQDGGGGLMCQENRNLWLLAGVTSFGYQCALNRRFGVAVRPR 1025
QY 845 FRDWIK 850
Db 1026 FTEWIK 1031

RESULT 4
A56318
enteropeptidase (EC 3.4.21.9) precursor [validated] - human
N:Alternate names: enterokinase
C:Species: Homo sapiens (man)
C:Date: 19-May-1995 #sequence_revision 09-Aug-1996 #text_change 09-Jul-2004
C:Accession: A56318; B43090
R:Kitamoto, Y.; Velle, R.A.; Donis-Keller, H.; Sadler, J.E.
Biochemistry 34, 4562-4568, 1995
A:Title: cDNA sequence and chromosomal localization of human enterokinase, the proteolytic
A:Reference number: A56318; MUID:95234679; PMID:7718557
A:Accession: A56318
A:Molecule type: mRNA
A:Residues: 1-1019 <KIT>
A:Cross-references: UNIPROT:P98073; GB:U09860; NID:G746412; PID:AAC50138.1; PID:G746413
R:Kitamoto, Y.; Yuan, X.; Wu, Q.; McCourt, D.W.; Sadler, J.E.
Proc. Natl. Acad. Sci. U.S.A. 91, 7588-7592, 1994
A:Title: Enterokinase, the initiator of intestinal digestion, is a mosaic protease compo
A:Reference number: A43090; MUID:94329561; PMID:8052624
A:Accession: B43090
A:Molecule type: mRNA
A:Status: nucleic acid sequence not shown
A:Residues: 749-1019 <KIT>
A:Cross-references: GB:U09860
A:Comment: The mechanism of association with the membrane of the intestinal brush border
oated below) or with amino-terminal myristoylation of the heavy chain.
C:Genetics:
A:Gene: GDB:PRSS7
A:Cross-references: GDB:384083; OMIM:226200
A:Map position: 21q21-21q21
C:Complex: Mature enteropeptidase is variously reported to contain two (heavy and light)
ed by a disulfide bond. Possibly, conversion from membrane-bound to soluble forms involv
ducts.
C:Function:
A:Description: cleaves activation peptide from trypsinogen to produce active trypsin
A:Pathway: intestinal digestive hydrolase cascade
C:Superfamily: enteropeptidase; Clr/Clis repeat homology; LDL receptor ligand-binding rep
C:Keywords: glycoprotein; hydrolase; serine proteinase; transmembrane protein; zymogen
F:1-784/Product: enteropeptidase heavy chain #status predicted <HCH>
F:22-38/Domain: transmembrane #status predicted <TMM>
F:184-221/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F:342-504/Domain: MAM homology <MAM>
F:526-631/Domain: Clr/Clis repeat homology <Clr>
F:643-677/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F:678-783/Domain: scavenger receptor cysteine-rich domain homology #status atypical <SRQ
F:785-1019/Product: enteropeptidase light chain #status predicted <LCH>
F:116,147,179,328,335,388,440,470,503,534,630,682,706,725,848,887,909,949/Binding site:
F:772-896,810-826,910-977,941-956,967-995/Disulfide bonds: #status predicted
F:825,876,971/Active site: His, Asp, Ser #status predicted

Query Match 14.6%; Score 682.5; DB 1; Length 1019;
Best Local Similarity 32.5%; Pred. No. 1.1e-36;
Matches 180; Conservative 80; Mismatches 202; Indels 91; Gaps 20;

QY 333 QLPRMSSCGG--RLRAQQTFFNSPPYGHYPNIDCTNIEVPNNQHVKVSFKFYLLP 390
Db 519 ELP--TDCGPELWEPNTTFSSTNFPNSYNPLAFVCWILNAQKGNITQLHPQEPDEN- 575
QY 391 GVPAGTCPKDYVEI-NGEK-----YCGERSQFVWTSNKNITVRFHSDQSYTDGFL 441

Db 576 -----INDVVEIRDGEADSLLLAVYTGPGVKDVFSTTNRMTVLLITNDVLARGGPK 628
QY 442 AEYLSS---YDSSDPCPG-QFTCRTRCIRKELRCGWDADCTDHSDELNCS CDAGHOF--- 494
Db 629 ANFTTGHGLPBECKADHFOCKNGECVPLVNLCDGHLHCEGSDGSDAECV-----RFPNG 683
QY 495 TCKN-----KPKCPKLPFWCDSVNDGDSDEOGSCSPAOTFRCSNGKCLSKSQOCCNGKDDC 550
Db 684 TTNNGLVRFRIQSIW-----HTACAEN-----WTTQISNDVC-----QLL 719
QY 551 GDSDEASCPKVNVTCTKHTYRCINGLCLSKGNPECDG-----KEDCSDGS-----D 598
Db 720 GLGNSNKKPFST-----DGOPFVKLNTAPDGHLLITPSQQLQSLRLRQCN 768
QY 599 EKDCDCGLRSFTRQARVVGTTDADEGEWPMQVSLHALQGHICGASLSPNWLVSAAHCY 658
Db 769 HKSCGKLAADITPKIVGGSSNAKEGAWPVPVGLY-YGRLLCGASLVSSDWLVSAAHCV 827
QY 659 IDRGFRYSPTQWTAFLGLHDQSQRSAPGVQERRLKRIISHPPFNDFTFYDIALLELE 718
Db 828 YG-----RNLEPSKWTAILGLHMKSNLTSPTQVPRPLIDEIVINPHYNNRRKKNDIAMHLE 883
QY 719 KPAEYSSMVRPICLPDASHVPPAGKAIWVTGWTGHTQVGTGALILQKGEIRVINQTTCE 778
Db 884 FKVNTDYIQICLPEENQVFPGRNCSIAWGTVVYQGTANTILQEAADVPLLSNERCQ 943
QY 779 LLPO-QITPRMNCVGLSGVDSQCGDGGGGLSSVEADGRIFQAGVVSWGDCGAQRNKP 837
Db 944 QMPEYNITENMICAGYEEGGIDSCGDSGGGLMCQENNRWFLAGVTSFGYKCALPNRPG 1002
QY 838 VYTRPLPLFRDWIK 850
Db 1003 VYARVSRFTEWIK 1015

RESULT 5
JE0315
low-density lipoprotein receptor-related protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 09-Jul-2004
C:Accession: JE0315
R:Tomita, Y.; Kim, D.; Magoori, K.; Fujino, T.; Yamamoto, T.T.
J. Biochem. 124, 784-789, 1998
A:Title: A novel low-density lipoprotein receptor-related protein with type II membran
A:Reference number: JE0315; MUID:98429596; PMID:9756624
A:Accession: JE0315
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1113 <TOM>
A:Cross-references: UNIPROT:Q92319; DBJ:AB013874; NID:93869144; PID:BAJ34371.1; PID:
C:Superfamily: mouse low-density lipoprotein receptor-related protein; LDL receptor 11
F:337-372/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F:374-408/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F:410-445/Domain: LDL receptor ligand-binding repeat homology <LDL3>
F:447-482/Domain: LDL receptor ligand-binding repeat homology <LDL4>
F:648-682/Domain: LDL receptor ligand-binding repeat homology <LDL5>
F:784-720/Domain: LDL receptor ligand-binding repeat homology <LDL6>
F:723-757/Domain: LDL receptor ligand-binding repeat homology <LDL7>
F:869-1097/Domain: trypsin homology <TRY>

Query Match 14.2%; Score 663.5; DB 2; Length 1113;
Best Local Similarity 32.7%; Pred. No. 2e-35;
Matches 158; Conservative 67; Mismatches 187; Indels 71; Gaps 16;

QY 428 RFHSDQSYTDGFLAEVLSYDSSDPCPGQFTRCIRKELRCGWDADCTDHSDELNCS 487
Db 627 QFPESDQNTCLLP---NEDVSCSPSHFKCRSGRCVLSGRCDGQADQDDSDENCG 683
QY 488 CDAGHQFTCK-NKPKRPLFWCDSVNDGDSNDGQCS-CPAQTFRCSNGKCLSKSQOCCN 545
Db 684 CKERALWECPPNKQCKLHTLICDGFDPDPSMDKNCSCFQDNELECANECVPRDLWCD 743
QY 546 GKDCGDSDEASCPKN-----VVTCTKHTYRCING-----LC----- 579

Db 744 GWVDCSDSDWECVTLTKXNGNSSLLTVHKSAREHVCADGWRETLSLQACKQOMGLGEP 803
QY 580 -LSKGNPECDGKE-----DCSDGSDK-----C---DCGLRST 610
Db 804 SVTKLPGQEGQWLLPYNWNENLNGSTQLLELVYRHSRSEISLLSKQDCGRPA 863
QY 611 R-QARVVGTDADGEPWPQVSLHALGQGHICGASLISPNWLSAAHCYIDDRGRFRYSDP 669
Db 864 RMNKRILGRTSPGRWPWQCSLQSPSGHICGVLLAKKWLTVACFP-EGR-----EDA 918
QY 670 TOWTAPLGLHDSQORSAPGVQERLKRILSHFPNDFDIDYDIALLELKPARYSMWRP 729
Db 919 DWKVVVFGNNLPHPSG-FMOTRFVKTILLHPRYSRAVVDYDISVVELSDDINETSYVRP 977
QY 730 ICLPDASHVPPKAKLWTVGWGHTVGGTICALILOKEIRVINQTTCEMLLPQO-ITPRM 788
Db 978 VCLPSPPEEYLEDPTYCIYIYNGH--MKNKMPKLQEGEVRIIPLOQCQSYFDMKTIITRM 1035
QY 789 MCVGLSGVDSCQDSDGGLSSVEADGRIFQAGVWVGDCGAQR-NKPGVYTRLEPLFRD 847
Db 1036 ICAGYESGTVDSCMGDSGGLVCPERPQGWTLFGLTSVGSVCFKVLGRGVISNVSYFVG 1095
QY 848 WIK 850
Db 1096 WIE 1098
RESULT 6
T30337
polyprotein - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C:Accession: T30337
R:Yang, J.C.; Lindeay, L.L.; Hedrick, J.L.
submitted to the EMBL Data Library, March 1998
A:Description: cDNA cloning of ovocytinase, a chymotrypsin-like protease released from X
A:Reference number: Z20829
A:Accession: T30337
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1524 <YAN>
A:Cross-references: UNIPROT:O91674; EMBL:U01290; NID:G2981640; PID:G2981641; PIDN:AAC247
C:Superfamily: trypsin related polypeptin; trypsin homology
Query Match 12.4%; Score 578.5; DB 2; Length 1524;
Best Local Similarity 26.2%; Pred. No. 1e-29;
Matches 167; Conservative 89; Mismatches 205; Indels 177; Gaps 23;
QY 238 YPAHARQWALRGDADSVLSLTFRSFDL---ASCDERGSGLVTVYNTILS-PMEPHALVOL 293
Db 333 YSINSVCRWMLAVQAKTIEIRFLQDIEDHATC-----TPDYLSFTVNEKMIRKV 383
QY 294 CGTPPSNLTFSHSONVLLITLNTERRHPGFATFPQ-P-RMSSCG-GRLRKQAGT 350
Db 384 CGSTIPSLIV---RSNKVTVTFFSDGTFTGRGFEIQFLAIPTKAASACGSAKILKKGM 440
QY 351 FNSPYPCGHPENIDCTWNIENPNHVKFRRFYLLEPGVPAG-TCPKDYVEI-NG-- 406
Db 441 IYSNPYDPYPRILKTCWIIIEAPENHIVKLKPEDF-----NVEYGHGCIYDAVEYVDGAE 495
QY 407 -----EKYCGERSQFVWTSNKNITVRPHSPQSYTDTGTFLAELIYSDSSDPCPGQFTCRT 461
Db 496 EKQILARLCGVTLPLPISSPENTMLIRPKTDMENSYPGFKVF----- 538
QY 462 GRCIRKELRCDGWADCTDHSBELNCSADAGHQFTCKNK-FCKPLFWVCDSDVNDGDNDSDE 520
Db 539 -----SFVPEKQFSLPV----- 551
QY 521 QGCSCPAQTFRCSNGKCLSKSQCCNGKDDCGDSDSEASCPCVNVVTCVTKHYRCINGLCL 580
Db 552 -----DD-----TPTISML----- 560

QY 581 SKGNPECDKEDCSDGSDKDCDCGLRSTFTRQ---ARVVGVTDADEGEFWQVSLHALGQ 637
Db 561 ---HPRALTD-----VCGMAPMTPKWLPRIVGGEEASPSNPWQVQIFFLT 606
QY 638 GHI CGASLISPNWLSAAHCYIDDRGRFRYSDPTQWTAFLGLHDSQORSAPGVQERLKR 697
Db 607 FH-CBGAIISQWILTAHC-----IRAAEFSYTVIAGDHNRLNNEST-EQIRNIKTI 658
QY 698 LSHFPNDFDIDYDIALLELKPARYSMWRPCLPDASHVPPKAKLWTVGWGHTVQYCG 757
Db 659 RIHDNYSNSETYDNDIALYLEPLDNDLPVPCVLPPEPEEVLT PASVCVTVTGWNTARDG 718
QY 758 TICALILOKEIRVINQTTCE-NLLPQQITPRMCMCVGLSG-GVDSQDSDGGLSSVEAD 815
Db 719 QPALGLQLQLPILDSIICNTSYYSSELTDHMLCAGFPSSKEKDACQDSDGGLVCQNEK 778
QY 816 GRIFQAGVWVGDCGAQRNKPGVYTRLEPLFRDWIKENT 853
Db 779 EQFSIYGLVSGEGCGRVSKPGVYTKVRLFFTWI-QNT 815
RESULT 7
154763
Ra-reactive factor (EC 3.4.21.-) 1 precursor - human
N:Alternate names: mannose binding protein-associated serine proteinase 1 (MASP-1)
C:Species: Homo sapiens (man)
C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 09-Jul-2004
C:Accession: 154763; JN0883
R:Sato, T.; Endo, Y.; Matsushita, M.; Fujita, T.
Int. Immunol. 6, 665-669, 1994
A:Title: Molecular characterization of a novel serine protease involved in activation
A:Reference number: 154763; MUID:94289349; PMID:8018603
A:Accession: 154763
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-699 <SAT>
A:Cross-references: UNIPROT:P48740; GB:D28593; NID:G790963; PIDN:BAA05928.1; PID:G4711
R:Takada, F.; Takayama, Y.; Hatause, H.; Kawakami, M.
Biochem. Biophys. Res. Commun. 196, 1003-1009, 1993
A:Title: A new member of the C1s family of complement proteins found in a bactericidal
A:Reference number: JN0883; MUID:94059062; PMID:8240317
A:Accession: JN0883
A:Molecule type: mRNA
A:Residues: 1-234; 'E', 236-284, 'G', 286-498, 'K', 500-542, 'K', 544-642, 'S', 644-699 <TAK>
A:Cross-references: DDBJ:DJ17525; NID:G439712; PIDN:BAA04477.1; PID:G439713
A:Experimental source: liver
C:Comment: This is a serum bactericidal factor that activates complement C4 and C2 com
C:Genetics:
A:Gene: GDB:MSP1; GDB:CRARF; CRARF1; PRSS5; MASP
A:Cross-references: GDB:361104; GDB:330954; OMIM:600521
A:Map position: 3q27-3q28
C:Superfamily: complement-activating serine proteases C1r/C1s/MASP; C1r/C1s repeat hom
C:Keywords: beta-hydroxyasparagine; complement pathway; duplication; glycoprotein; hyd
F:1-17/Domain: signal sequence #status predicted <SIG>
F:18-448, 449-699/Product: Ra-reactive factor #status predicted <MAT>
F:19-135/Domain: C1r/C1s repeat homology <C1R1>
F:143-181/Domain: EGF homology <EGF>
F:185-294/Domain: C1r/C1s repeat homology <C1R2>
F:301-362/Domain: complement factor H repeat homology <FHL>
F:367-432/Domain: complement factor H repeat homology <FHD>
F:449-691/Domain: trypsin homology <TRY>
F:49, 178, 407/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:73-91, 143-157, 153-166, 168-181, 185-212, 242-260, 301-349, 329-362, 367-414, 397-432, 436-57
F:159/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted
F:448-449/Cleavage site: Arg-Ile (autolytic) #status predicted
F:490, 552, 646/Active site: His, Asp, Ser #status predicted
Query Match 12.3%; Score 574; DB 1; Length 699;
Best Local Similarity 25.3%; Pred. No. 8.4e-30;
Matches 193; Conservative 111; Mismatches 268; Indels 192; Gaps 32;
QY 203 SKTVQRTQDNCSCFGLHARGVELMRFPTTPGPPSPYPAHARQWALRGDADSVLSLTFRS 262

Db 15 SKASATVLENNMFG-----QIOSFGYDPS-YPSDSEVTWNITVPDGFRIKLYFMH 64
QY 263 FOLASCDERGSIDLVTYNTLSPMEPHALVOLCG-----TYPPSY-NLTFH 306
Db 65 FNELESSYLCEYDAVKV-----ETEDQVLATFCGRETITDTEQTPGQEVVLSFGSFWMSITFR 119
QY 307 SSONVLLIITLITNTERHFGFEATFPQLP----- 335
Db 120 SD-----FENEERFTGDAHYMAVDVDECKEREDBELSCDHYCHNYIGGYCSCRF 170
QY 336 -----RMSSCGRLRAQGFNSPYYPGHYPNIDCTWNLEVENNQHVKSFPK 384
Db 171 GYLILHNDNTRCECDNLFTQRTGVITSPDFNFPKSECELYTELEBEGFVNLQFED 230
QY 385 FYLLE--PGVPAGTCKPYVEIN-GEK-----XGERSQFVVTNSNKKITVRHFSQSYTD 437
Db 231 IPDIQDHPEVP--CPYDIKIKVGPVKLPGFCGEKAPETISTQSHSVLILFHSNDAEN 287
QY 438 TQFLABYLSYDSDDCP-----GOFTCRGCRKELRCGWADCTDHSDELNCSDA 490
Db 288 RGWRLSYRA--AGNECEPELQPPVHGKIEPSQAKYFPK-----DQVLVSCDT 331
QY 491 GHOF-----YCKNFKCKPLFW-----VCDVNDGDNDSBQSCCPAQ-----TFR 531
Db 332 GYKVLKDNVEMDTFQIECLKDGTMWNKIPTCKIVD-----CRAPGELEHGLITFS 381
QY 532 CSNGKCLSKSQNGKXDCGDSDEASCPKVNVTCTKHTYRC-----LNLGLCLSKGNP 585
Db 382 TRNLTYYSEI--KYSQE-----PYRKLNNNTGIYCSAQQVMWNV-LGRSLP 430
QY 586 ECDGKEDCSDGSEKDCDGLRSFTQ--ARVVGTDADGEMWQVLSLHALGQGHICGA 643
Db 431 TCLPV-----CGLPKFSRKLMAFINGRPAQKGTTPMTAMLSHLNGOPFCGG 477
QY 644 SLISPNWLSAHCY--ID--DRGRYSD--PTOWTAFGLHDQSRASGVQERRLK 695
Db 478 SLIGSWMIVTAHLQHSIDGPTLRDSDLSPSKFIILGKH-WRLSDSNEHGLGVK 536
QY 696 RIISHPFNDFDYDIALLELEKPAEYSSMVRPCLPDASHVFPAGKAIWVTGHTQY 755
Db 537 HTTLHPQYDNTFENDVALVELLESPLVNAFWPCLPESGPQ--QEGAMVIVSGWG-KQF 593
QY 756 GTGALILQGEIRVINQTCENL---LPQQTTPRMVCVGFSLSGVDSQCGSGGSLSV 812
Db 594 LORFPETLMEIPIVDHSTCKAVAPLKKVTRDMICAGEKEGKDACAGDSGGPMVTL 653
QY 813 EAD-GRIFOAGVYVSMGDCGAQRNKGVTYTRLPFRDWIKENTGV 855
Db 654 NREGQWYLVGVVSMGDDCGKDRYGVYSIIHNNKDWIORVTGV 697

RESULT 8

QKMSPL
plasma kallikrein (EC 3.4.21.34) precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 09-Jul-2004
C:Accession: A36557
R:Seidah, N.G.; Sawyer, N.; Hamelin, J.; Mion, P.; Beaubien, G.; Brachpapa, L.; Rochemond
DNA Cell Biol. 9, 737-748, 1990
A:Title: Mouse plasma kallikrein: cDNA structure, enzyme characterization, and comparison
A:Reference number: A36557 MUID:91090844; PMID:2264928
A:Accession: A36557
A:Molecule type: mRNA
A:Residues: 1-638 <SEI>
A:Cross-references: UNIPROT:P26262; GB:M58588; NID:G200358; PID:AAA3393.1; PID:G200359
A:Note: part of this sequence, including the amino ends of both the heavy and light chain
C:Comment: This protein, synthesized in the liver, circulates as a noncovalent complex w
C:Comment: The zymogen is activated by factor XIIa, which cleaves the molecule into a li
are linked by one or more disulfide bonds.
C:Superfamily: coagulation factor XI; trypsin homology
C:Keywords: blood coagulation; duplication; fibrinolysis; glycoprotein; hydrolase; infla
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-390/Product: plasma kallikrein heavy chain #status experimental <HCH>

F:20-109/Domain: apple repeat <AP1>
F:110-199/Domain: apple repeat <AP2>
F:200-289/Domain: apple repeat <AP3>
F:291-380/Domain: apple repeat <AP4>
F:391-638/Product: plasma kallikrein light chain #status experimental <LCH>
F:391-621/Domain: trypsin homology <TRY>
F:121-104,47-77,51-57,111-194,137-166,141-147,201-284,227-256,231-237,292-375,318-347,3
F:127,215,308,396,494/Binding site: carbohydrate (Aen) (covalent) #status predicted
F:434,483,578/Active site: His, Asp, Ser #status predicted

Query Match 11.4%; Score 533.5; DB 1; Length 638;
Best Local Similarity 27.1%; Pred. No. 3.3e-27;
Matches 183; Conservative 86; Mismatches 233; Indels 173; Gaps 31;
QY 271 RGSLLVTVNTLSPMEPHALVOLCGTYPSPSYNLTFHSSQNVLITLITNTERHFGF--E 328
Db 32 RGGDLAAIYTP-----DAQYCKMCTFHPCLLFSF-----LAVTPKETNRKFCFMKE 81
QY 329 ATFFQLPRMSSCG--GELRKAQGFNSPYYPGHYPF-----NIDCTWNIE-----V 372
Db 82 SITGTLPRIHRTGAISHLSKQCGHQISACHRDYIKGLDMRGSNFIKTDNIESCQKLC 141
QY 373 PNNQHVK--VSFKFFYLLE-----PGVPAQT-----C 397
Db 142 TNHFHCQFTVATSAFTRPEYRKCLLKHSASTPTSIKSADNLVSGFSLKSCALSEIGC 201
QY 398 PKD-----YVEINGEKYCGERS--QFVVTNSNKKITVRHFSQSYTDG----FLABY 444
Db 202 PMDIFQHSAPADNLVSVQVITPDADFVCRITCFHFENCLFFTFYTNWETESQRNVCLKTS 261
QY 445 LSYDSSDPCPQ-----FTCRGTR--CIRKELRCDGWADCTDHSDELNCSQDAGH- 492
Db 262 KSGRPSPIPOENAIISGVSLTCTCRTPPEPCHSKI-----YSGVDFFEGLNVTVOGAD 316
QY 493 --QFTCKNFKCKPLFWCDSV-NDGDNDSQSCGCPAQTPRCNSNGKCLSKSQCGCKDD 549
Db 317 VCQETC-TKTIRCQFFIYSLPQDC---KEEGKC---SURLST----- 353
QY 550 CGDGSDEASCPKVNVTCTKHTYRLNG-----LCLSKGNPECDGKEDCDGSDKDCD 603
Db 354 --DGSP-----TRITYGQSGSYSLAKCLVDSPTCTKIN----- 388
QY 604 CGLSFTFQARVVGTDADGEMWQVLSLHA--LGQGHICGASLISPNWLSAHCYIDD 661
Db 389 -----ARIVGTTNAGLGEWPMQVSLQKLVSTHLCCGSIIGROWMLVTAACHF--- 436
QY 662 RGRYSDPTQWTAFLGLHDQSQ--RSAPGVQERLKRISHPFFNDFDYDIALLELEK 719
Db 437 DGIPY--PDVWRIYGGILSLSEITKTP--SSRIKELIIHQEYKVSSENYDIALIKQT 491
QY 720 PAEYSSMVRPCLPDASHVFPAGKAIWVTGHTQYGGTGALILQKGEIRVINQTCENL 779
Db 492 PLNVTEFQKPICLPSKADNTIYTNWVTGWTGVTKEQGETQNILQKATIRLPVNECQK 551
QY 780 LPQO-ITPRMVCVFLSGVDSQCGDGGPLSSVEADGRIFOAGVVSNGDCCGAQRNKG 838
Db 552 YRDVINKQMTICAGYKGGTDACKDGGGPL-VCKHSGRWQLVGTISWEGCGCKRQDQGV 610
QY 839 YTRLPLFRDWIKENT 853
Db 611 YTKVSEYMDWILEKT 625

RESULT 9

KORTPL
plasma kallikrein (EC 3.4.21.34) precursor - rat
N:Alternate names: Fletcher factor; kininogenin; serum kallikrein
C:Species: Rattus norvegicus (Norway rat)
C:Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 09-Jul-2004
C:Accession: A39180; S06851; I53041; S06852
R:Beaubien, G.; Rosinski-Chupin, I.; Mattei, M.G.; Mbikay, M.; Chretien, M.; Seidah, N
Biochemistry 30, 1628-1635, 1991
A:Title: Gene structure and chromosomal localization of plasma kallikrein.

A;Reference number: A39180; MUID:91129236; PMID:1993180
A;Accession: A39180
A;Molecule type: DNA
A;Residues: 1-638 <BEA>
A;Cross-references: UNIPROT:P14272; GB:J05315
A;Note: the authors translated the codon GAG for residue 81 as Gln
R;Seidah, N.G.; Ladenheim, R.; Molkay, M.; Hamelin, J.; Lutfalla, G.; Rougeon, F.; Lazou
DNA 8, 563-574, 1989
A;Title: The cDNA structure of rat plasma kallikrein.
A;Reference number: A33320; MUID:90091743; PMID:2598771
A;Accession: A33320
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-638 <SEI>
A;Cross-references: GB:M30282; NID:G205010; PIDN:AAA41463.1; PID:G205011
A;Note: part of this sequence, including the amino ends of both the heavy and light chain
R;Paquin, J.; Benjannet, S.; Sawyer, N.; Lazure, C.; Chretien, M.; Seidah, N.G.
Biochim. Biophys. Acta 999, 103-110, 1999
A;Title: Rat plasma kallikrein: purification, NH(2)-terminal sequencing and development
A;Reference number: S06851; MUID:90089457; PMID:2597701
A;Accession: S06851
A;Molecule type: protein
A;Residues: 20-45;391-413 <PAQ>
R;Seidah, N.G.; Ladenheim, R.; Molkay, M.; Hamelin, J.; Lutfalla, G.; Rougeon, R.; Lazou
DNA Cell Biol. 8, 563-574, 1989
A;Title: The cDNA structure of rat plasma kallikrein.
A;Reference number: I53041
A;Accession: I53041
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-638 <RES>
A;Cross-references: GB:M58590; NID:G206721; PIDN:AAA42069.1; PID:G206722
A;Comment: This protein, synthesized in the liver, circulates as a noncovalent complex
C;Comment: The zymogen is activated by factor XIIa, which cleaves the molecule into a 11
are linked by one or more disulfide bonds.
C;Genetics:
A;Gene: PK
C;Superfamily: coagulation factor XI; trypsin homology
C;Keywords: blood coagulation; duplication; fibrinolysis; glycoprotein; hydrolase; infla
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-390/Product: plasma kallikrein heavy chain #status experimental <MAT1>
F;20-109/Domain: apple repeat <AP1>
F;110-195/Domain: apple repeat <AP2>
F;200-285/Domain: apple repeat <AP3>
F;291-380/Domain: apple repeat <AP4>
F;391-638/Product: plasma kallikrein light chain #status experimental <MAT2>
F;391-621/Domain: trypsin homology <TRY>
F;21-104,47-77,151-57,111-194,137-166,141-147,201-284,227-256,231-237,292-375,318-347,322
F;127,215,308,453,459,494/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;396/Binding site: carbohydrate (Asn) (covalent) #status experimental
F;434,483,578/Active site: His, Asp, Ser #status predicted
Query Match 11.0%; Score 514.5; DB 1; Length 638;
Best Local Similarity 26.5%; Pred. No. 5.8e-26;
Matches 183; Conservative 81; Mismatches 218; Indels 209; Gaps 35;
Qy 271 RGSLLVTVYNTLSPEPHALVOLCTYPPSYNLTTFHSSQNVLLITNTERRHGF--E 328
Db 32 RGGDLAAYI---TPDAQHC-QKMTFHPRCLLFSP-----LAVSPTKETDRRCFWKE 81
Qy 329 ATFFQLPRMSSCG---GRLRKAQGFNSPYYPGHYPP-----NIDCTWNIE-----V 372
Db 82 SITGTLPRHRTGATSGHSLKQCGHLSACHQDIYEGLDMRGNSFNISKTDISECQKLC 141
Qy 373 PNNQHVK---VSEKFFYLE-----PGVPAGTCPKDYVEING--EKYCG----- 411
Db 142 TNNIHCQFFYATKAFHPEYRKSCLLRSSSGTPTSIPKVDNL-VSGFSLKSCALSBEIG 200
Qy 412 -----ERSQFVTSNSKNTVRHSDQSYTDTGTFLAELVSDSDPCPGOFTCTGRCI 465
Db 201 CPMDFIHFADLVNSQVLT-----PDAFVCKT----- 229
Qy 466 KRELRCDGWADCTDHSDELNGSCDAGHQTCKNFKCKPLF-----WVCDSV-NDCGDN 517

Db 230 -----VCTFHP---NC-----LFFTYTNWETESQRNVCFLK 259
Qy 518 SDEQCSCPA-----QTFRCSNGKCLSKSQCGKDCG-----DGS 554
Db 260 TSKGRSPFPITQENAVSGYSLETCRK-----ARPECFKFIYSGVAFEGEELNATFVOGA 315
Qy 555 D--EASCPKVVVTCVTKYRCLNGLCLSKGNPECDGKEDCS-----DGS----- 597
Db 316 DACQETCTK--TIRCQFTYSLLPDCKAEG--C-----KCSLRSLTDSGSPRITVEAOG 366
Qy 598 -----DEKDCDGLRSFTRQARVVGGTDADEGEWPQVSLHA--LQGHICGA 643
Db 367 SSGYSLRLCKVVESSDCTTKI-----NARIVGGTNSLSGEWQVSLQVLSQNHMCGG 421
Qy 644 SLISPNLWLSAAHCYIDDRGFYSPTOWTAFGLHDQSO--RSAPGVQERLRKLIISHP 701
Db 422 SIIGQWILTTAAHCF--DGIPY--PDVWRIYGGILNLSEITNKTTP--FSSIKSLIHO 473
Qy 702 PFNDFTFDYDIALLEKPAESYSSMVRPICLPDASHVFPAGKAIWVTGHTQYGGTGAL 761
Db 474 KYMSEGSYDIALIKLQTPLYNTEFFQKICLPKSRADTNTIYNCVWVGTYKGETON 533
Qy 762 ILQKGEIRVINQTTCCNLLPQO--ITPRMVCVGLSGVDSCGDSGGPLSSVEADRIPO 820
Db 534 ILQKATIPLVNPEECQKRYDVWITKQICAGYKEGGIDACKGDSGGPL-VCKHSGRWOL 592
Qy 821 AGVYSGDGGCAQRNKGYYTFLPLFROWIKE 851
Db 593 VGIWSGEGCARKQPGVYTKVAEYIDWILE 623
RESULT 10
PLPG
plasmin (EC 3.4.21.7) precursor - pig (fragment)
N;Alternate names: plasminogen
N;Contains: miniplasminogen
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 07-Sep-1990 #sequence revision 01-Nov-1996 #text_change 09-Jul-2004
C;Accession: S03733; S03737; A25834
R;Schaller, J.; Marti, T.; Roesselet, S.J.; Kaempfer, U.; Rickli, E.E.
Fibrinolysis 1, 91-102, 1987
A;Title: Amino acid sequence of the heavy chain of porcine plasmin. Comparison of the
A;Reference number: S03733
A;Accession: S03733
A;Molecule type: protein
A;Residues: 1-560 <SCH>
R;Brunisholz, R.A.; Lerch, P.G.; Schaller, J.; Rickli, E.E.; Lergier, W.; Manneberg, M
Eur. J. Biochem. 114, 465-470, 1981
A;Title: Comparison of the primary structure of the N-terminal CNBr fragments of human
A;Reference number: S03735; MUID:81212097; PMID:7238497
A;Accession: S03737
A;Molecule type: protein
A;Residues: 1-57 <BRU>
R;Marti, T.; Schaller, J.; Rickli, E.E.
Eur. J. Biochem. 149, 279-285, 1985
A;Title: Determination of the complete amino-acid sequence of porcine miniplasminogen.
A;Reference number: A25834; MUID:85203907; PMID:3846533
A;Accession: A25834
A;Molecule type: protein
A;Residues: 450-790 <MAR>
C;Function:
A;Description: dissolves the fibrin of blood clots; acts as a proteolytic factor in a
ns the walls of the graafian follicle; also activates the urokinase-type plasminogen a
A;Pathway: fibrinolysis
C;Superfamily: plasmin; kringle homology; plasminogen-related protein precursor homolo
C;Keywords: fibrinolysis; glycoprotein; hydrolase; kidney; kringle; plasma; serine pro
F;1-790/Product: plasminogen #status predicted <PRO>
F;1-77/Domain: plasminogen-related protein precursor homology (fragment) <PLPH>
F;1-77/Domain: activation peptide #status predicted <APT>
F;78-560/Product: plasmin chain A #status predicted <ACH>
F;84-162/Domain: kringle homology <KRI>

Db 756 GVTSWGLGCLALPNKPGVYVRSRFTWIEE 785

RESULT 11

S15468

complement C3b/C4b inactivator (EC 3.4.21.-) precursor - African clawed frog

N;Alternate names: C3b/C4b inactivator factor I

C;Species: Xenopus laevis (African clawed frog)

C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 09-Jul-2004

C;Accession: I51601; S15468

R;Kunnath-Muglia, L.M.; Chang, G.H.; Sim, R.B.; Day, A.J.; Ezekowitz, R.A.

Mol. Immunol. 30, 1249-1256, 1993

A;Title: Characterization of Xenopus laevis complement factor I structure--conservative

A;Reference number: I51601; MUID:94019415; PMID:7692240

A;Accession: I51601

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-613 <KUN>

A;Cross-references: UNIPROT:Q03711; EMBL:X59958; NID:G64595; PIDN:CA442582.1; PID:G645

C;Superfamily: human complement factor I; LDL receptor ligand-binding repeat homology;

C;Keywords: hydrolase; serine proteinase

F;220-254/Domain: LDL receptor ligand-binding repeat homology <LDL1>

F;257-290/Domain: LDL receptor ligand-binding repeat homology <LDL2>

F;370-599/Domain: trypsin homology <TRY>

Query Match 10.8%; Score 506; DB 2; Length 613;

Best Local Similarity 26.2%; Pred. No. 2e-25;

Matches 153; Conservative 81; Mismatches 181; Indels 168; Gaps 22;

QY 397 CPKD---YVEINGEK-----YCGERSQFVTSNSKNIYRPHSDQSYDTGTGFLAEVLSYDS 449

Db |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||

65 CPKNATTEVCTDGRKLOQSYCOLKS--VECSNPLNSKYRFPSEAPCTETFTLTQ-----N 117

QY 450 SDP-----CPQOFTCRGCRIRKELRCGWDACT----- 478

Db |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||

118 GFPGKGIIVKVLPTFPEGLFCGQNSRANVVCRLGSGTKGADASADKVFSLVTEKP 177

QY 479 -DHSDELNC-----SC-----DAGH-QFTCKNFKCKPLFW 506

Db |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||

178 PEHCIQATCRGLENSLAECALRKLPMQDNQVAKVTCYTENKDCGGEFTCSNGKCIPESEL 237

QY 507 VCDSDVNDGNSDEQGC-SCPQATFRCSNGKCLSKSOCCNKDDCGDGSDEASC----- 559

Db |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||

238 ACDSKNDGDLSDLCCKSCNA-GFHCRSPTCIPEQYRCNGELDCIGGEDESNCTVBEQ 596

QY 560 -----PKNVVTCTKHTYRCLNGCLCLKSGNPECD 588

Db |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||

297 KSEKQBEVEQKQTSKEQEBDLVQESKATQVEEKAINVYDIDAERL---LMSLPE--- 350

QY 589 GKEDCDSDGDEKDCDCL----RSFTQARVVGGTDADGEWPMQVSLHALGOGHI CG 642

Db |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||

351 -----LSCGVPPQTAALTILTKRKRVIGGINAVKNQFPQWVAIKD-GTAVNCG 396

QY 643 ASLISPNLWLSAAHYIDRGRFYSDDPTQWTAFLGL-----HDGQRGAPGVQERRLKRI 697

Db |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||

397 GIYIGGCWLTAAHC-----VRSNQQRVYLIMLELLDRLSYDKOLDSPF-----VKSV 444

QY 698 ISHPFENDFTFDYDIALLEL-----KPAEYSMSVRPICLPDASHVPPAKAIWVTGWG 751

Db |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||

445 IVHELNPNTYENDIALLEVIYNNPKWADNNVWPACVPWSPFQFAGTCTVSGWG 504

QY 752 HTQYGGTGALILQKGEIRVINOTTTCENLLPQOITPRMMCVGFLSGGVDS CQDGSQGPLSS 811

Db |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||

505 R-EKGSRSRVPHLKWGHINLMD--NCTRVYKERFLDKMECARTYDGSIDACKDGSQGPLVC 561

QY 812 VADGRIFQAGVYSGWDGCAQRNKPQYVTRLPFRDWIKENTG 854

Db |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||

562 YDNKVAYVWGIWSGNGCVPVGPVYTKVYANIFEWIAHQVG 604

RESULT 12

B61545

plasmaIn (EC 3.4.21.7) precursor - sheep (fragments)


```

Query Match      10.6%; Score 497; DB 1; Length 810;
Best Local Similarity 24.6%; Pred. No. 1.1e-24;
Matches 234; Conservative 104; Mismatches 294; Indels 288; Gaps 49;

QY      102 VDAENSNSTEVSLSKVKDALKLLYSGVPFLGPGYHKESAVTAFSGSVTAYYWSF-- 159
          : : : : : : : : : : : : : : : : : : : : : : : : : :
Db       22 LDDYNTQGSLSFVTKQ-----LGAGSIECAAKCED-----ESFTC 61
          : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      160 -SIPQHLVEEASRVMAEERVVMLPFRARSLSKSVFTTSVVAFFTDSK-----TVQRT 209
          : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

Db 62 RAFQHSKEQQCVIMAEKRSKSIIRMDV---VLFKVKVYLSECKTGNKNGYRGTMSKT 118
QY 210 QDNSSCSFGLHARGVELMRFTT-----PGFPDGPYPAHA---RCO-----WALRGD 251
Db 119 KN-----GITCKWSSSPHPRSPATHSEGLEENYCRNPNDPOGPWCYTTD 168
QY 252 ADSVLSLTPRSPDLASCDER-----GSDL-VTVYNTLSPE-----PHALVOLCGTY 297
Db 169 PEK---RYDYCDILECEBECMHCSGENYDGKISKMTSGLECOAWDSQSPA---HGYI 220
QY 298 P---PSYNTLTPHSSQNVLITITWTERP-----HPGEATFFQOLPRMS-----SCGRL 344
Db 221 PSKFNKNLKNYCRN-----PDRELAPWCFTTDPNKRWELCDIPRCTTTPPPSSGPTY 273
QY 345 RKAQGTENSPPYYPGHYPNIDCT-----WNIEVPNNQHVKSFFFYLLLEPGVPAGT 396
Db 274 QCLKGTE-----NYRGNVAVTVSGHTCOHWSAQTPHTN-----R 309
QY 397 CPKQVVEIN-GEKIC-----GERSQFVVTNSNKKITVRHSDQSYTDTGFLAPYLSYDSSD 451
Db 310 TPEFPCKNLDENYCRNPDGKAPWCHTTNSQ---VRW-----EYCKIPSCD 353
QY 452 PCP-----CO-----FTCRGRCIRKELRCDCGWADCTDH 480
Db 354 SSPVSTEQLAPTELPVQVQDYGHDQSYRGTSSTTTGK-----KQSWSSMTPH 407
QY 481 SDELNCSCDAGHQFTCKNKFKFLFWVCDVNDGDNDSDEQCSC---PAQTER-CSNG 535
Db 408 R-----HOKTPENYPNAGL-----TMNYCRNPADKGPWCFTTDPSPVRWEYCNLK 452
QY 536 KC-----LSKSQOCNGKDDC--GDG-----SDEAS----- 558
Db 453 KCSGTEASVAVPPVLLPNVETPSEEDCMFGNGKGYRGKRATTVTGTPCODWAAQEPHR 512
QY 559 ----CPKVNVTCTKHTYRCLN--GLCLSKGNPEC--DGKEDCSGSEKDC-----DCG 605
Db 513 HSIPTETNPRAGLEKNY-CRNPDG---DVGGPWCYTTNPRKLYDYCDVPQCAAPSFDCG 568
QY 606 LRSFTRQ---ARVVGTDADGEWQVSLHALGQGHICGASLISPNWLVSAAHCIYDDR 662
Db 569 KPQVEPKKCPGRVUGGVCAHPSHWPQVSLRTRFGMHFCGGTLISPEWVLTAAHCL--EK 626
QY 663 GFRYSDPTQWTAFLGLHDQSQRSAPGVOERRLKRIISHPFFNDFTFDYDIALLELEKPAE 722
Db 627 SPR---PSSYKVLGAH-QEVNLEPHVQIEVSRLEPFRK-----DIALKLSSPAV 676
QY 723 YSSMRPDIPLDASHVFPAGKAIWTVGHTQYGGTGALILQKGEIRVINOTTCE--NLL 780
Db 677 ITDKVIPACLPSPNTVAVDRTECFITGWGETQ-GTFGAGLLKEAQLPVLENKVCNRYEFL 735
QY 781 PQOITPRMVCVFLSGSDSCGSDGGLSSVEADGRIFOAGVVSWMGDGCAORNKPGVYT 840
Db 736 NGRVOSTELCAGLAGGTDSCGSDGGLPVCFEKDKYILQ-GVTSWGLGCAARNKPGVYV 794
QY 841 RLPLFRDWIK 850
Db 795 RVSREVTWIE 804

```

Search completed: September 23, 2005, 12:56:06
Job time : 30 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 23, 2005, 12:49:05 ; Search time 179 Seconds
(without alignments)
2445.965 Million cell updates/sec

Title: US-09-421-213-2

Perfect score: 4681

Sequence: 1 MGSDRARKGGGPKDFAGL.....PGVYTRFLPRDWIKENTGV 855

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt_O3:*

1: uniprot_spot:*

2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4676	99.9	855	1 ST14 HUMAN	Q9Y5Y6 homo sapien
2	3901	83.3	855	1 ST14 MOUSE	P56677 mus musculus
3	3883	83.0	855	2 Q9JJI7	Q9JJI7 rattus norv
4	2690	57.5	845	2 Q6GR54	Q6GR54 xenopus lae
5	2664	56.9	845	2 Q9DGR1	Q9DGR1 xenopus lae
6	2653	56.7	845	2 Q63ZQ6	Q63ZQ6 xenopus lae
7	2379	50.8	422	2 Q8WVC1	Q8WVC1 homo sapien
8	1969	42.1	663	2 Q6DEV0	Q6DEV0 xenopus tro
9	1124.5	24.0	799	2 Q6PF94	Q6PF94 mus musculus
10	1124.5	24.0	811	1 TMS6_MOUSE	Q8db10 mus musculus
11	1124	24.0	811	1 TMS6_HUMAN	Q8lu80 homo sapien
12	1122	24.0	802	2 Q6UXD8	Q6uxd8 homo sapien
13	1102	23.5	824	2 Q6ICC2	Q6iccc2 homo sapien
14	1011.5	21.6	572	2 Q7RTY8	Q7rtys8 homo sapien
15	1011.5	21.6	572	2 Q8BIK6	Q8bik6 mus musculus
16	727	15.5	1034	1 ENTK_PIG	P98074 sus scrofa
17	717.5	15.3	855	2 Q7Z410	Q7z410 homo sapien
18	717.5	15.3	1059	2 Q7Z411	Q7z411 homo sapien
19	712.5	15.2	1035	1 ENTK_BOVIN	P98072 bos taurus
20	692	14.8	1042	1 COR1_HUMAN	Q9Y5G5 homo sapien
21	690.5	14.8	1111	2 Q80YN4	Q80yn4 rattus norv
22	687	14.7	777	2 Q8CAN9	Q8can9 mus musculus
23	682.5	14.6	1019	1 ENTK_HUMAN	P98073 homo sapien
24	676.5	14.5	1069	1 ENTK_MOUSE	P97435 mus musculus
25	663.5	14.2	1113	1 COR1_MOUSE	Q9Z319 mus musculus
26	644.5	13.8	767	2 Q9DGR2	Q9dgr2 xenopus lae
27	644	13.8	722	2 Q6NUF5	Q6nuf5 xenopus lae
28	636.5	13.6	680	2 Q868H7	Q868h7 branchiosto
29	623.5	13.3	680	2 Q868H5	Q868h5 branchiosto
30	617	13.2	581	2 Q9XZM7	Q9xzm7 strongyloce
31	612	13.1	688	2 Q868H6	Q868h6 branchiosto

RESULT 1

ST14_HUMAN	601	12.8	490	2	Q7TN04	Q7tn04 mus musculus
AC Q9Y5Y6; Q9BS01; Q9H3S0; Q9HB36; Q9HCA3; PRT; 855 AA.	600	12.8	490	1	TMS2_MOUSE	Q9jiq8 mus musculus
DT 16-OCT-2001 (Rel. 40, Created)	600	12.8	490	2	Q920K3	Q920k3 rattus norv
DT 25-OCT-2004 (Rel. 45, Last sequence update)	598.5	12.8	490	2	Q6P7D7	Q6p7d7 rattus norv
DE Suppressor of tumorigenicity 14 (EC 3.4.21.-) (Matrptase) (Membrane-type serine protease 1) (MT-SP1) (Prostamin) (Serine protease TADG-15) (tumor associated differentially-expressed gene-15 protein).	597	12.8	688	2	Q868H4	Q868h4 branchiosto
GN Name=ST14; Synonyms=PRSS14, SNC19, TADG15; Homo sapiens (Human).	593.5	12.7	701	2	Q9JJS9	Q9jijs9 rattus norv
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	588	12.6	422	1	DES1_HUMAN	Q9ul52 homo sapien
OX NCBI_TaxID=9606;	587.5	12.6	1134	2	Q6UW31	Q6uw31 homo sapien
RN [1]	587	12.5	703	2	Q7RTY7	Q7rtty7 homo sapien
RP SEQUENCE FROM N.A.	585.5	12.5	704	1	CRAR_MOUSE	Q8chn8 rattus norv
RX MEDLINE=99303581; PubMed=10373424; DOI=10.1074/jbc.274.25.18231; Lin C.Y., Anders J., Johnson M., Sang Q.A., Dickson R.B.;	585.5	12.5	868	2	Q91IV3	P98064 mus musculus
RA "Molecular cloning of cDNA for matriptase, a matrix-degrading serine protease with trypsin-like activity.";	578.5	12.4	1524	2	Q91674	Q91674 xenopus lae
RT J. Biol. Chem. 274:18231-18236(1999).	577	12.3	676	2	Q6DUJ6	Q6duj6 cyprinus ca
RL [2]						
RN [3]						
RP SEQUENCE FROM N.A.						
RX MEDLINE=99432178; PubMed=10500122; DOI=10.1073/pnas.96.20.11054; Takeuchi T., Shuman M.A., Craik C.S.;						
RA "Reverse biochemistry: use of macromolecular protease inhibitors to dissect complex biological processes and identify a membrane-type serine protease in epithelial cancer and normal tissue.";						
RL Proc. Natl. Acad. Sci. U.S.A. 96:11054-11061(1999).						
RN [4]						
RP SEQUENCE FROM N.A.						
RC TISSUE=Prostate;						
RA Yamaguchi N., Mitsu S.;						
RT "Molecular cloning of a novel transmembrane serine protease expressed in human prostate.";						
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.						
RN [5]						
RP SEQUENCE FROM N.A.						
RA Tanimoto H., Underwood L.J., Wang Y., Shigemasa K., Parnley T.H., O'Brien T.J.;						
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.						
RN [6]						
RP SEQUENCE FROM N.A.						
RC TISSUE=Blood, and Muscle;						
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899; Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,						
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,						
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,						
RA Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.D., Scheetz T.E., Brownstein M.J., Uadin T.B., Toshiyuki S., Carninci P., Prange C.,						
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,						

ALIGNMENTS

RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Ketterman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shervenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RL [6]
RP SEQUENCE OF 340-664 FROM N.A.
RA Cao J., Fan W., Zheng S.;
RT "Genomic analysis of a novel human serine protease SNCl9";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RL [7]
RN CHARACTERIZATION.
RP TISSUE=Milk;
RX MEDLINE=99303582; PubMed=10373425; DOI=10.1074/jbc.274.26.18237;
RA Lin C.Y., Anders J., Johnson M., Dickson R.B.;
RT "Purification and characterization of a complex containing matrixase
RT and a Kunitz-type serine protease inhibitor from human milk";
RL J. Biol. Chem. 274:18237-18242(1999).
CC -!- FUNCTION: Degrades extracellular matrix. Proposed to play a role
CC in breast cancer invasion and metastasis. Exhibits trypsin-like
CC activity as defined by cleavage of synthetic substrates with Arg
CC or Lys as the P1 site.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
CC -!- SIMILARITY: Belongs to the peptidase S1 family.
CC -!- SIMILARITY: Contains 2 CUB domains.
CC -!- SIMILARITY: Contains 4 LDL-receptor class A domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF118224; AAD4765.2; -;
CC EMBL; AF133086; AAF00109.1; -;
CC EMBL; BAB03036; BAB20376.1; -;
CC EMBL; AF057145; AAG15395.1; -;
CC EMBL; BC005826; AAH05826.2; -;
CC EMBL; BC030532; AAH05832.1; -;
CC EMBL; AF283256; AAG13949.1; -;
CC PDB; 1EAW; X-ray; A/C#615-855.
CC MEROPS; S01.302; -;
CC Genew; HGNC:11344; Srt14.
CC MIM; 606797; -;
CC GO; GO:0005887; C:integral to plasma membrane; TAS.
CC GO; GO:0008236; F:serine-type peptidase activity; TAS.
CC GO; GO:0006508; P:proteolysis and peptidolysis; TAS.
CC InterPro; IPR000859; CUB.
CC InterPro; IPR002172; LDL receptor A.
CC InterPro; IPR009003; Pept Ser_Cys_1.
CC InterPro; IPR001254; Peptidase_S1.
CC InterPro; IPR001314; Peptidase_S1A.
CC Pfam; PF00431; CUB; 2;
CC Pfam; PF00057; Ldl_recept_a; 4.
CC Pfam; PF00089; Trypsin; 1;
CC PRINTS; PR00722; CHYMOTRYPSIN.
CC PRINTS; PR00261; LDLRECEPTOR.
CC SMART; SM00042; CUB; 2;
CC SMART; SM00192; LDLA; 4.
CC SMART; SM00020; Tryp_SPC; 1.
CC PROSITE; PS01180; CUB; 2;
CC PROSITE; PS01209; LDLRA_1; 2;
CC PROSITE; PS00668; LDLRA_2; 4;
CC PROSITE; PS0240; TRYPSIN_DOM; 1.

DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW 3D-structure; Glycoprotein; Hydrolase; Repeat; Serine protease;
KW Signal-anchor; Transmembrane.
FT DOMAIN 1 55 Cytoplasmic (Potential).
FT TRANSMEM 56 76 Signal-anchor for type II membrane
FT protein (Potential).
FT CUB 1. Extracellular (Potential).
FT CUB 2.
FT DOMAIN 214 334 LDL-receptor class A 1.
FT DOMAIN 340 447 LDL-receptor class A 2.
FT DOMAIN 452 487 LDL-receptor class A 3.
FT DOMAIN 487 524 LDL-receptor class A 4.
FT DOMAIN 524 560 Serine protease.
FT DOMAIN 566 603 Charge relay system (By similarity).
FT DOMAIN 615 854 Charge relay system (By similarity).
FT ACT_SITE 711 711 Charge relay system (By similarity).
FT ACT_SITE 805 805 Charge relay system (By similarity).
FT CARBOHYD 109 109 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 302 302 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 485 485 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 772 772 N-linked (GlcNAc...) (Potential).
FT CONFLICT 381 381 R -> S (in Ref. 4).
FT CONFLICT 674 674 A -> V (in Ref. 3).
SQ SEQUENCE 855 AA; 94769 MW; 26143132C01F99C9 CRC64;
Query Match 99.9%; Score 4676; DB 1; Length 855;
Best Local Similarity 99.9%; Pred. No. 1.2e-312;
Matches 854; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MGSRRARKGKGGPKDFAGLKNRSHKRVNGLSEGVFLPVNNVKKVKGKGGPRWVLA 60
DB 1 MGSRRARKGKGGPKDFAGLKNRSHKRVNGLSEGVFLPVNNVKKVKGKGGPRWVLA 60
QY 61 VLIGLLVLLGIGLFWHLQYRDVRQKVFNGVWRIITNENFVDAYENSNTFVSLASKV 120
DB 61 VLIGLLVLLGIGLFWHLQYRDVRQKVFNGVWRIITNENFVDAYENSNTFVSLASKV 120
QY 121 KDALKLYSGVPFLGPKYHKSATVAFSEGSVIAYYWFSEFSIPQHLVEBARVMAERVVM 180
DB 121 KDALKLYSGVPFLGPKYHKSATVAFSEGSVIAYYWFSEFSIPQHLVEBARVMAERVVM 180
QY 181 LPPRARSLSFVTSVVAFTDSTKVQRTQNSCSFGLHARGVELMRTTFGFPDPSPPA 240
DB 181 LPPRARSLSFVTSVVAFTDSTKVQRTQNSCSFGLHARGVELMRTTFGFPDPSPPA 240
QY 241 HARCQWALRGDADSVLSITFRSPDLASCDERGSDLVTVYNTLSPMEPHALVOLCCTYPPS 300
DB 241 HARCQWALRGDADSVLSITFRSPDLASCDERGSDLVTVYNTLSPMEPHALVOLCCTYPPS 300
QY 301 YNLTFHSSQNVLLITLTINTERRHPGFATPFQPLRMSSCGGLRKAKGTFFNSPYPGHY 360
DB 301 YNLTFHSSQNVLLITLTINTERRHPGFATPFQPLRMSSCGGLRKAKGTFFNSPYPGHY 360
QY 361 PPNTDCTWNIETVNNQHVKSFKFYLLEPGVPAGTCKDYVEINGEKYQGERQFVYTS 420
DB 361 PPNTDCTWNIETVNNQHVKSFKFYLLEPGVPAGTCKDYVEINGEKYQGERQFVYTS 420
QY 421 NSNKITVRFHSDQSYTDTGFLAEYLSYDSDPCPGQFTCTGRCIRKELRCDGWDCTDH 480
DB 421 NSNKITVRFHSDQSYTDTGFLAEYLSYDSDPCPGQFTCTGRCIRKELRCDGWDCTDH 480
QY 481 SDELNCSADAGHQFTCKNFKCKPLFWVCDSDVNDGSDNSDEQGCSCPAQTRCNSGKCLSK 540
DB 481 SDELNCSADAGHQFTCKNFKCKPLFWVCDSDVNDGSDNSDEQGCSCPAQTRCNSGKCLSK 540
QY 541 SQQCNKGKDDCGDSDSEASCPRKVNVTCTKTYRCINGLCLSKGNPECDGKDCSDGSDSK 600
DB 541 SQQCNKGKDDCGDSDSEASCPRKVNVTCTKTYRCINGLCLSKGNPECDGKDCSDGSDSK 600
QY 601 DCDGLRSFTQARVVGTDADGEWPMQVSLHALGQGHICGASLISPNMLVSAHCVYID 660
DB 601 DCDGLRSFTQARVVGTDADGEWPMQVSLHALGQGHICGASLISPNMLVSAHCVYID 660

QY 661 DRGFRYSPTQWTAFLGLHDQSQRAPGVQERLKRILSHHPFNFTFDYDIALLELEKP 720
 Db 661 DRGFRYSPTQWTAFLGLHDQSQRAPGVQERLKRILSHHPFNFTFDYDIALLELEKP 720
 QY 721 AYSWSVRPCLPDASHVFPAGKAIWVTGWGHTQYGGTICALIQLGEIRVINQTCENLL 780
 Db 721 AYSWSVRPCLPDASHVFPAGKAIWVTGWGHTQYGGTICALIQLGEIRVINQTCENLL 780
 QY 781 PQOITPRMVCVFLSGGVDSQCGDGGPLSSVEADGRIFQAGVSWGDCQARNKPGVYT 840
 Db 781 PQOITPRMVCVFLSGGVDSQCGDGGPLSSVEADGRIFQAGVSWGDCQARNKPGVYT 840
 QY 841 RLPLFRDWIKENTGV 855
 Db 841 RLPLFRDWIKENTGV 855

RESULT 2

ST14 MOUSE
 ID ST14_MOUSE STANDARD; PRT; 855 AA.
 AC P56677;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 25-OCT-2004 (Rel. 45, Last annotation update)
 DE Suppressor of tumorigenicity 14 (SC 3.4.21.-) (Epithin).
 GN Name=St14; Synonyms=Prs814;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C.B.17SCID; TISSUE=Thymus;
 RX MEDLINE=9216440; PubMed=1019918; DOI=10.1007/s002510050515;
 RA Kim M.G., Chen C., Lyl M.S., Cho E.G., Park D., Kozak C.,
 RA Schwartz R.H.;
 RT "Cloning and chromosomal mapping of a gene isolated from thymic
 RT stromal cells encoding a new mouse type II membrane serine protease,
 RT epithin, containing four LDL receptor modules and two CUB domains.";
 RL Immunogenetics 49:420-428(1999).
 RN [2]
 RP REVISIONS TO 23; 321; 325; 343; 409-410 AND C-TERMINUS.
 RC STRAIN=C.B.17SCID; TISSUE=Thymus;
 RA Kim M.G., Chen C., Cho E.G., Park D., Schwartz R.H.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Breast tumor;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Wang J.J., Heide F.,
 RA Datschenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huilyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J.J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzyzanski M.I., Skalska U., Smallos D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
 CC -1- TISSUE SPECIFICITY: Highly expressed in intestine, kidney, lung,
 CC and thymus. Not expressed in skeletal muscle, liver, heart, testis

CC and brain.
 CC -1- SIMILARITY: Belongs to the peptidase S1 family.
 CC -1- SIMILARITY: Contains 2 CUB domains.
 CC -1- SIMILARITY: Contains 4 LDL-receptor class A domains.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF042822; AAD02230.3; -;
 CC EMBL; BC005496; AAH05496.1; -;
 CC HSSP; P00760; LEZX.
 CC MEROPS; S01.302; -;
 CC MGD; MGI:1338881; St14.
 CC GO; GO:0005576; C:extracellular; IDA.
 CC GO; GO:0019897; C:extrinsic to plasma membrane; IDA.
 CC GO; GO:0008236; F:serine-type peptidase activity; IDA.
 CC InterPro; IPR000859; CUB.
 CC InterPro; IPR002172; LDL_receptor_A.
 CC InterPro; IPR009003; Pept_Ser_Cys.
 CC InterPro; IPR001254; Peptidase_S1.
 CC InterPro; IPR001314; Peptidase_S1A.
 CC Pfam; PF00431; CUB; 2.
 CC Pfam; PF00057; Ldl_recept_a; 4.
 CC Pfam; PF00089; Trypsin; 1.
 CC PRINTS; PR00722; CHYMOTRYPSIN.
 CC PRINTS; PR00261; LDLRECEPTOR.
 CC SMART; SM00042; CUB; 2.
 CC SMART; SM00192; LDLA; 4.
 CC SMART; SM00020; Tryp_SPC; 1.
 CC PROSITE; PS01180; CUB; 2.
 CC PROSITE; PS01209; LDLRA_1; 2.
 CC PROSITE; PS00068; LDLRA_2; 4.
 CC PROSITE; PS02040; TRYPSIN_DOM; 1.
 CC PROSITE; PS00134; TRYPSIN_HIS; 1.
 CC PROSITE; PS00135; TRYPSIN_SER; 1.
 KW Glycoprotein; Hydrolase; Repeat; Serine protease; Signal-anchor;
 KW Transmembrane.
 FT DOMAIN 1 55 Cytoplasmic (Potential).
 FT TRANSMEM 56 76 Signal-anchor for type II membrane
 FT protein (Potential).
 FT DOMAIN 77 855 Extracellular (Potential).
 FT DOMAIN 214 331 CUB 1.
 FT DOMAIN 340 444 CUB 2.
 FT DOMAIN 451 488 LDL-receptor class A 1.
 FT DOMAIN 489 522 LDL-receptor class A 2.
 FT DOMAIN 523 561 LDL-receptor class A 3.
 FT DOMAIN 565 604 LDL-receptor class A 4.
 FT DOMAIN 615 854 Serine protease.
 FT ACT_SITE 656 656 Charge relay system (By similarity).
 FT ACT_SITE 711 711 Charge relay system (By similarity).
 FT ACT_SITE 805 805 Charge relay system (By similarity).
 FT CARBOHYD 107 107 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 302 302 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 365 365 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 421 421 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 489 489 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 772 772 N-linked (GlcNAc...) (Potential).
 SQ SEQUENCE 855 AA; 94654 MW; 4F10B84DA2146DD5 CRC64;
 Query Match 83.3%; Score 3901; DB 1; Length 855;
 Best Local Similarity 81.8%; Pred. No. 2,1e-259;
 Matches 699; Conservative 73; Mismatches 83; Indels 0; Gaps 0;
 QY 1 MGSDRARKGGGKDFGAGLKYNSRHEKYNKNGEGVEFLPVNNVKVEKHGPGRWVLLA 60
 Db 1 MGSNRKRGKGGSQDFGAGLKYNSRLENNMGFEVEFLPVNNVKVEKHGPGRWVLLA 60
 QY 61 VLIGLLLVLLGIGFLVWHLQYRDVVRQKVFNGYWRITNENFVDAYENSNSTEFVSLASKV 120

```
Db 61 VLFSFLLSLMAGLLVWHFYHNRVQKVFNGHLRTNFIPLDAYENSTSTFEFISLASQV 120
Qy 121 KDALKLLYGVPLPGPYHKESAVTAFSEGSVLAAYWSEFIPQHLVEEAERVMABERVVM 180
Db 121 KEALKLLYNEVPLGPHKKSAVTAFSEGSVLAAYWSEFIPPHLAEVDRAVERVVT 180
Qy 181 LPPRARSLKSFVVTSSVVAFTDSTKVORTQDSCSFGHARGVELMRFTTPGFPDPSYPYA 240
Db 181 LPPRARALKSFVLTSSVVAFTDPRMLORTQDSCSFPALHAAGAAVTRFTTPGFPNPSYPYA 240
Qy 241 HARCQWALRGDADSVLSLTFRSFDLASCDERGSDLVTYNTLSPMEPHALVOLCGTYPPS 300
Db 241 HARCQWVLRGDADSVLSLTFRSFDVAPCDHEGSDLVTVYDLSLSPMEPHAVVRLCGTFSPS 300
Qy 301 YNLTFHSSQNVLLITLITNTERRHPGFEATFFQLPKMSCCGRLKAKOCTFNSPYYPGHY 360
Db 301 YNLTFSSQNVFLVTLITNTDRRHGPGFEATFFQLPKMSCCGFLSDTQGTGTFSPYPGHY 360
Qy 361 PPNIDCTWNIETVNNQHVKSFKFYLLLEPGVPAGTCTPKDYVEINGEKYCGERSQFVVTs 420
Db 361 PPNINCTWNIKVPNNRNKVRKFLFYLVDPNVPGSCTKDYVEINGEKYCGERSQFVVSs 420
Qy 421 NSNKITVRHSDQSYTDTGFLAEYLSYDSDRCPGQFTCTGRCIRKELRCGDWADCTDH 480
Db 421 NSSKITVPHFSHRSYTDGFLAEYLSYDSDNCPGGMFMCKTGRKIRKELRCGDWADCPDY 480
Qy 481 SDELNCSCDAGHQFTCKNFKCKPLFWVCDVNDGDNDSDEQSCSPAOFTFRCSNGKCLSK 540
Db 481 SDERYCRNATHQFTCKNQFCPLFWVCDVNDGDSDEQSCSPAGSKCSNGKCLLPQ 540
Qy 541 SQCKNGKDCGDSDEASCPKVVNVYCTKHYTRCLNGLCLSKNGNPECDKEDCSDGSDX 600
Db 541 SQCKNGKONCGDSDEASCDVNVVSCYKYTVRCNGLCLSKNGNPECDKEDCSDGSDX 600
Qy 601 CDCGLRSFTRQARVVGTTDADGEPWQVSLHALGQGHICGASLISPNLVSAAHCYID 660
Db 601 NCDCGLRSFTRQARVVGTTADGEPWQVSLHALGQGHICGASLISPNLVSAAHCYID 660
Qy 661 DRGFRYSDPTQWTAFLGLHDQSRAPGQVQERRKRIISHPPFNFTFDYDIALLELEKP 720
Db 661 DKNFKYSDYTWMTAFGLLDQSKRASGVQELKRLIITHPSFNFTFDYDIALLELEKS 720
Qy 721 ABYSNWRPICIPLDASHVPPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTTCEMLL 780
Db 721 VEYSTWRPICIPLDATHVPPAGKAIWVTGWGHTKEGGTGCALILQKGEIRVINQTTCEML 780
Qy 781 PQOITPRMVCVFLSGGSDSCGDSGGLSSVEADGRIFQAGVSWGDCQQRNKPQYIT 840
Db 781 PQOITPRMVCVFLSGGSDSCGDSGGLSSAEKDGMRMFQAGVSWGEGCAQRNKPQYIT 840
Qy 841 RLPLFRDWIKENTGV 855
Db 841 RLPLVRDWIKENTGV 855
ID Q9JUI7 PRELIMINARY; PRT; 855 AA.
AC Q9JUI7;
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DT 05-JUL-2004 (TremBLrel. 27, Last annotation update)
DE Membrane bound serine protease (Membrane bound arginine specific
DE serine protease).
GN Name=MSP;
OC Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=wistar; TISSUE=Jejunum;
```

```
RX MEDLINE=21459307; PubMed=11573963; DOI=10.1006/bbrc.2001.5686;
RA Satomi S., Yamasaki Y., Tezuka S., Hitomi Y., Iwanaga T., Fushiki T.;
RT "A role for membrane-type serine protease (MT-Sp1) in intestinal
RL epithelial turnover";
RN Biochem. Biophys. Res. Commun. 287:995-1002(2001).
RP [2]
SEQUENCE FROM N.A.
RC STRAIN=wistar; TISSUE=Duodenum;
RA Inoue H., Takahashi K., Kishi K.;
RL Submitted (SSP-2000) to the EMBL/GenBank/DBJ databases.
CC -I- SIMILARITY: Belongs to peptidase family S1.
DR EMBL; AB037898; BAB03502.1; -
DR EMBL; AB049189; BAB13765.1; -
DR PIR; JC7731; JC7731.
DR HSSP; Q9Y5Y6; IEAW.
DR MEROPS; S01.302; -.
DR GO; GO:0004263; P:chymotrypsin activity; IEA.
DR GO; GO:0008233; P:peptidase activity; IEA.
DR GO; GO:0004235; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR001172; LDL receptor A.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_SIA.
DR InterPro; IPR009003; Pept_Ser_Cys.
DR Pfam; PF00431; CUB; 2.
DR Pfam; PF00057; Ldl_recept_a; 4.
DR Pfam; PF00089; Trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00261; LDLRECEPTOR.
DR PROSITE; PS01180; CUB; 2.
DR PROSITE; PS01209; LDLRA_1; 2.
DR PROSITE; PS00068; LDLRA_2; 4.
DR PROSITE; PS02040; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR Hydrolase; Serine protease.
KW Hydrolyase; Serine protease.
SQ SEQUENCE 855 AA; 94955 MW; 35806B7BCF6CF03D CRC64;
Query Match 83.0%; Score 3883; DB 2; Length 855;
Best Local Similarity 81.1%; Pred. No. 3,6e-258;
Matches 693; Conservative 79; Mismatches 83; Indels 0; Gaps 0;
Qy 1 MGSDRARKGGGKPKDPCGAGLKYNRSHKVNLEGEVFLPVNNVKKYKKGPGRWVLA 60
Db 1 MGNRRGKAGGSDGDFGAGLKYNRSLNMGNGFEGVEFLPVNNAKQVEKRGPRRVMVA 60
Qy 61 VLIGLLVLLGIGPLVWHLOYRDVVRQKVFNGYMRITNENFVDAYENSSTFEFVLSASKV 120
Db 61 VVFSFLLSLMAGLLVWHFYHNRVRIQKVFNGHLRTNENFPLDAYENSTFEFISLASQV 120
Qy 121 KDALKLLYGVPLPGPYHKESAVTAFSEGSVLAAYWSEFIPQHLVEEAERVMABERVVM 180
Db 121 KEALKLLYNEVPLGPHKKSAVTAFSEGSVLAAYWSEFIPPHLAEVDRAVERVVT 180
Qy 181 LPPRARSLKSFVVTSSVVAFTDSTKVORTQDSCSFGHARGVELMRFTTPGFPDPSYPYA 240
Db 181 LPPRARALKSFVLTSSVVAFTDPRMLORTQDSCSFPALHAAGAAVTRFTTPGFPNPSYPYA 240
Qy 241 HARCQWALRGDADSVLSLTFRSFDLASCDERGSDLVTYNTLSPMEPHALVOLCGTYPPS 300
Db 241 HARCQWVLRGDADSVLSLTFRSFDVAPCDGHDSDLVTVYDLSLSPMEPHAVVRLCGTFSPS 300
Qy 301 YNLTFHSSQNVLLITLITNTERRHPGFEATFFQLPKMSCCGRLKAKOCTFNSPYYPGHY 360
Db 301 YNLTFSSQNVFLVTLITNTDRRHGPGFEATFFQLPKMSCCGFLSDTQGTGTFSPYPGHY 360
Qy 361 PPNIDCTWNIETVNNQHVKSFKFYLLLEPGVPAGTCTPKDYVEINGEKYCGERSQFVVTs 420
Db 361 PPNINCTWNIKVPNNRNKVRKFLFYLVDPNVPGSCTKDYVEINGEKYCGERSQFVVSs 420
Qy 421 NSNKITVRHSDQSYTDTGFLAEYLSYDSDRCPGQFTCTGRCIRKELRCGDWADCTDH 480
```


QY	491	GHQFTCKN-KFCXPLFWVCDVNDGNSDEQSCCPAOTFRCSNGKCLSKSQCCNGKDD	549
Db	479	-LQFRCVNSKLCXPSYFICDGVNDGSDSDCLACKCPNNTFKCGNKCIIPDSQKCDRVN	537
QY	550	CGDGSDEACPKVNVVCTKHTYRCLNGLCLSGNPECCKEDCSGDSDEK--DCDCGLR	607
Db	538	CGDGSDEACDQVLTACTEYTKCKNNOCITKKNPECDDGENDCSGSDENAACNCGKR	597
QY	608	SPTQARVVGTTDADGEWPMQVSLHALGQCHTCGASLISPNWLVSAAHCYIDDRGFYS	667
Db	598	PFTKSRIVGVGNADTGEFPWQVSLHAKGNKHTCGASLSVPTMLISAACHFODDHQMRIS	657
QY	668	DPTQWTAFLGLHDQSQRSAPGVQERRLKRIISHPFNDFTDYDIALLEKPAEYSNV	727
Db	658	DASLWTAYLGLHDQAQLNTKDVVERIKRINAHIGFNDNTYDNDIAVLELEKPEVYTDPI	717
QY	728	RPICLPDASHVFPAGKAIWVTGWGHTQYGGTGALLIOLKEIRVINQTTCCENLLPOQITPR	787
Db	718	QPVCIPSTHDFPVGKPIWVTGWGALKKGGGAAVILQKAEIRIINQTECNKLLDGLTPR	777
QY	788	MMCVGLSGVDSCQDGSGLSSVEADGRIFOAGVVSMDGCAORNKPGVYTRLPPLRD	847
Db	778	MLCAGFVSGGIDACQDGSGLSSVELNNKVVLAGVVSNGEGCARRNKEGVITKVSMMRD	837
QY	848	WIKENTGV	855
Db	838	WSKDTGL	845
RESULT 5			
Q9DGR1			
ID	Q9DGR1	PRELIMINARY; PRT; 845 AA.	
AC	Q9DGR1		
DT	01-MAR-2001	(TrEMBLrel. 16, Created)	
DT	01-MAR-2001	(TrEMBLrel. 16, Last sequence update)	
DT	01-MAR-2004	(TrEMBLrel. 26, Last annotation update)	
DE	Homolog of human MT-SPI.		
GN	Names=MT-SPI.		
OS	Xenopus laevis (African clawed frog).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;		
OC	Xenopodinae; Xenopus.		
OX	NCBI_TaxID=8355;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE-20363741; PubMed-10903452; DOI=10.1016/S0378-1119(00)00225-0;		
RA	Yamada K., Takabatake T., Takeshima K.;		
RT	"Isolation and characterization of three novel serine protease genes		
RT	from Xenopus laevis."		
RL	Gene 252:209-216(2000).		
CC	-1- SIMILARITY: Belongs to peptidase family S1.		
DR	EMBL; AB038498; BAB03218.1; -;		
DR	HSSP; Q9Y516; IEAW.		
DR	GO; GO:0004263; F:chymotrypsin activity; IEA.		
DR	GO; GO:0008233; F:peptidase activity; IEA.		
DR	GO; GO:0004295; F:trypsin activity; IEA.		
DR	GO; GO:0005508; P:proteolysis and peptidolysis; IEA.		
DR	InterPro; IPR000859; CUB.		
DR	InterPro; IPR002172; LDL receptor A.		
DR	InterPro; IPR001254; Peptidase S1.		
DR	InterPro; IPR001314; Peptidase S1a.		
DR	InterPro; IPR009003; Pept_Ser_Cys.		
DR	Pfam; PF00431; CUB; 2.		
DR	Pfam; PF00057; Ldl_recept_a; 4.		
DR	Pfam; PF00089; Trypsin; 1.		
DR	PRINTS; PR00722; CHYMOTRYPSIN.		
DR	PRINTS; PR00261; LDLRECEPTOR.		
DR	SMART; SM00042; CUB; 2.		
DR	SMART; SM00192; LDLa; 4.		
DR	SMART; SM00020; Tryp_SPC; 1.		
DR	PROSITE; PS01180; CUB; 2.		

DR	PROSITE; PS01209; LDLRA_1; 2.		
DR	PROSITE; PS00068; LDLRA_2; 4.		
DR	PROSITE; PS00240; TRYPSTN_DOM; 1.		
DR	PROSITE; PS00134; TRYPSTN_HIS; UNKNOWN_1.		
DR	PROSITE; PS00135; TRYPSTN_SER; 1.		
KW	Hydrolase; Protease; Serine protease.		
SQ	SEQUENCE 845 AA; 93597 MW; 7FD7E62851A758B7 CRC64;		
QY	14	KDFAGLKYNRHEKNGLEGEVFLPVNNVKVKEKHGPRVWVLAALVIGLLVLLGTG	73
Db	2	KDSMMKYNRRPQSLNGFEEGEVFLPATNSKKVEKTPKKKLAIFGLVIGALLSLTIG	61
QY	74	FLVHQLQVRDVRQVFNQVNRITNENFVDAYENSSTFEVSLASKVDAKALLVSGVFP	133
Db	62	LLVWHFAYRNKPVNKLTYGILTIANPTPFIDAYENSTTAEFSDLSAKVIDTLQTVYNGKD	121
QY	134	LGPYHKEAVTAFSEG---SVIAYYWSFSPHQLVVEAEERVMAEERVVMLEPPRARSLS	190
Db	122	IAPYLQKCSISAFSEGGNNVIGYWSFSPAPREAAFEKAISELKLPVSNPRQR---T	178
QY	191	FVVTSVVAFPTDSKTQRTDNCSFGLHARGVELMRFTTFCGPDSPYPAHARCOWALRG	250
Db	179	FALDSLVAYPTDPOIARVFKNSSCAYFLHSSNGVAKFSFGFDPSPYPRNARCLWTLRA	238
QY	251	DADSVLSLTPRSFDLASCDEGSDLVTVYNTLSPMEPHALVOLCGTYRPSYNLTFHSSQN	310
Db	239	DAGRIIHLHFKTQMEKCKPNGDFFVMVYDLSPIEPPRAQIRLCGIYPPSYNLTFFSSN	298
QY	311	VLLITLITNERRHPGEATFFQLPRMSSCGRLRKAQGTNPSYPYGHYPNIDCTWNI	370
Db	299	VMLVTLVTDNVGKPPGLAEFKQLPKTSLCGLLRDASGFTSTPFPYPAHYPSTESIMDI	358
QY	371	EVNNQHVVSFKFFYLEPGVPAGTCKPVVEINGEKYCGERSOFVVTSSNKTIVRPH	430
Db	359	QVPDNKFVKVRFNMFYLAEPGPVVTCTKDFEIKGQKYCGEKEFFVVSNNSSKMSVRFV	418
QY	431	SDQSVTDGFLAEVLSYDSDPCGCOFTCTGRGCIKELRCGDGADCTDHSDELNCSDA	490
Db	419	SDQSVTDTGTAELVLSYEPRNCPDQFTCRSGRCIRLQKCDGWNDCBDFDEMCCTVA	478
QY	491	GHQFTCKN-KFCXPLFWVCDVNDGNSDEQSCCPAOTFRCSNGKCLSKSQCCNGKDD	549
Db	479	-LQFRCVNSKLCXPSYFICDGVNDGSDSDCLACKCPNNTFKCGNKCIIPDSQKCDRVN	537
QY	550	CGDGSDEASCPKVVNVCTKHTYRCLNGLCLSGNPECCKEDCSGSDSK---DCDCGLR	607
Db	538	CGDGSDEACDQVLTACTEYTKCKNNOCITKKNPECDDGENDCSGSDENAAKNCCKR	597
QY	608	SPTQARVVGTTDADGEWPMQVSLHALGQCHTCGASLISPNWLVSAAHCYIDDRGFYS	667
Db	598	PFTKSRIVGVGNADTGEFPWQVSLHAKGNKHTCGASLGFTMLISAACHFODDHQMRYS	657
QY	668	DPTQWTAFLGLHDQSQRSAPGVQERRLKRIISHPFNDFTDYDIALLEKPAEYSVMV	727
Db	658	DASLWTAYLGLHDQAQLNTKDVVERIKRINAHIGFNDNTYDNDIAVLELEKPEVYTDPI	717
QY	728	RPICLPDASHVFPAGKAIWVTGWGHTQYGGTGALLIOLKEIRVINQTTCCENLLPOQITPR	787
Db	718	QPVCIPSTHDFPVGKPIWVTGWGALKKGGGAADVILQKAEIRIINQTECNKLLDGLTPR	777
QY	788	MMCVGLSGGVDSCQDGSGLPLSSVEADGRIFOAGVVSMDGCAORNKPGVYTRLPFLRD	847
Db	778	MLCAGFVSGGIDACQDGSGLPLSSVELNNKVVLAGVVSNGEGCARRNKEGVITKVSMMRD	837
QY	848	WIKENTGV	855
Db	838	WSKDKTGL	845

```

RESULT 6
Q63ZQ6 PRELIMINARY; PRT; 845 AA.
AC Q63ZQ6;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Hypothetical protein.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;
RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
RA Richardson P.;
RA "Genetic and genomic tools for Xenopus research: The NIH Xenopus
RT initiative.";
RL Dev. Dyn. 225:384-391 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Trinchard J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Goughwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smaluk D.E., Scherch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RL and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA Klein S., Gerhard D.S.;
RL Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC082854; AAH02854.1; -.
KW Hypothetical protein.
SQ SEQUENCE 845 AA; 93710 MW; 6B15E0E7BAA91A9 CRC64;

Query Match 56.7%; Score 2653; DB 2; Length 845;
Best Local Similarity 54.6%; Pred. No. 1.2e-173;
Matches 463; Conservative 153; Mismatches 22; Indels 10; Gaps 5;

Qy 14 KDFGGLKYNRHEKVGLEGVLEPVPNNVKKVKGHPGRVVLAAVLGLLLVLIG 73
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 74 FLVHVLQYRVKQVFNQYRNITNENFVDAYNSNSTEFYSLASKVDALKLYSGVPF 133
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 62 LLVWHFAFNAPTHKLYGYLTANTQIEAYNSTTFEADLAKVINTLQTVYNGKD 121
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 134 LQPYKHSAVTAREG-----SVIATYWFSEFSIPQHLVEAEARVMAERYVMLPPRARSILKS 190
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 122 IAPYLQCCSISARESEGNNGNVIYWFSEFVPAFREAAFEKAISEFKLPSPVDPRR---T 178
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 191 FVTSVVAFPDTSKTVQRTQNSCSFGLHARGVELMRTTTPGPDSPYAHARCWALRG 250
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 7
Q8WVC1 ID Q8WVC1 PRELIMINARY; PRT; 422 AA.
AC Q8WVC1;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE ST14 protein (Fragment).
GN Name=ST14;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Colon;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

```

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettaman M., Madan A.C., Shevchenko Y., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 [2]
 RN SEQUENCE FROM N.A.
 RP TISSUE=Colon;
 RA Strausberg R.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: Belongs to peptidase family S1.
 DR EMBL; BC018146; AAI18146.1; -;
 DR HSSP; Q9Y5Y6; IEAW.
 DR GO; GO:0004263; F:chymotrypsin activity; IEA.
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0004295; F:trypsin activity; IEA.
 DR GO; GO:000508; P:proteolysis and peptidolysis; IEA.
 DR Pfam; PF00057; Ldl_recept_a; 4.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR PRINTS; PR00261; LDLRECEPTOR.
 DR SMART; SM00192; LDLA; 4.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS01209; LDLRA_1; 2.
 DR PROSITE; PS50068; LDLRA_2; 4.
 DR PROSITE; PS0240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 KW Hydrolase; Protease; Serine protease.
 FT NON_TER 1
 SQ SEQUENCE 422 AA; 46257 MW; 2C99875D1858B319 CRC64;

Query Match 50.8%; Score 2379; DB 2; Length 422;
 Best Local Similarity 100.0%; Pred. No. 3.5e-155;
 Matches 422; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 434 SYTDGFLAEYLSYSSDPCPGQFTCTGRCIRKELRCGWDCTHSDLNCSDCAGHQ 493
 DB 1 SYTDGFLAEYLSYSSDPCPGQFTCTGRCIRKELRCGWDCTHSDLNCSDCAGHQ 60
 QY 494 FTCKNFKCKPLFWVCDSDVNDGSDSCPAOTFCRCNGKCLSKSQCCNGKDDCGD 553
 DB 61 FTCKNFKCKPLFWVCDSDVNDGSDSCPAOTFCRCNGKCLSKSQCCNGKDDCGD 120
 QY 554 SDEASCPKNVVTCTGTYRCLNGLCLSKGNPECDGKEDCSGDEKDCGLRFTTQQA 613
 DB 121 SDEASCPKNVVTCTGTYRCLNGLCLSKGNPECDGKEDCSGDEKDCGLRFTTQQA 180
 QY 614 RVVGGTDADEGEWQVSLHALGQGHICGASLISPNWLVSAAHCYIDDRGFYSPTQWT 673
 DB 181 RVVGGTDADEGEWQVSLHALGQGHICGASLISPNWLVSAAHCYIDDRGFYSPTQWT 240
 QY 674 AFLGLHDQSORANPGVQERLKRILISHPEFNDFTDYDIALLEKPAEYSSMVRPICLP 733
 DB 241 AFLGLHDQSORANPGVQERLKRILISHPEFNDFTDYDIALLEKPAEYSSMVRPICLP 300
 QY 734 DASHVFPAGKALVWTGWGHTQGGTGTGALILQGEIRVINTQTTCEMLLPQITPRMVCVF 793
 DB 301 DASHVFPAGKALVWTGWGHTQGGTGTGALILQGEIRVINTQTTCEMLLPQITPRMVCVF 360
 QY 794 LSGGVDSQCGDSGGLPSLSEADGRIFQAGVSWGDCGCAQRNKPQVYTRLPFRDWIKENT 853
 DB 361 LSGGVDSQCGDSGGLPSLSEADGRIFQAGVSWGDCGCAQRNKPQVYTRLPFRDWIKENT 420
 QY 854 GV 855
 DB 421 GV 422

RESULT 8

Q6DEV0 PRELIMINARY; PRT; 663 AA.
 ID Q6DEV0
 AC Q6DEV0
 DT 25-OCT-2004 (TrEMBLrel. 28, Created)
 DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
 DE MGC89623 protein.
 GN Name=MGC89623;
 OS Xenopus tropicalis (Western clawed frog) (Silurana tropicalis).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 OC Xenopodinae; Xenopus.
 OX NCBI_TaxID=8364;
 RP [1]
 RP SEQUENCE FROM N.A.
 RX TISSUE=Embryo;
 RX PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Uedin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raba S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettaman M., Madan A.C., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RA Klein S., Gerhard D.S.;
 RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC076994; AA076994.1; -;
 DR GO; GO:0004295; F:trypsin activity; IEA.
 DR GO; GO:000508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000859; CUB.
 DR InterPro; IPR00172; LDL_receptor_A.
 DR InterPro; IPR001254; Peptidase_S1.
 DR InterPro; IPR009003; Pept_ser_Cys.
 DR Pfam; PF00431; CUB; 2.
 DR Pfam; PF00057; Ldl_recept_a; 4.
 DR Pfam; PF00089; Trypsin; 1.
 DR PRINTS; PR00261; LDLRECEPTOR.
 DR SMART; SM00042; CUB; 2.
 DR SMART; SM00192; LDLA; 4.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS01180; CUB; 2.
 DR PROSITE; PS01209; LDLRA_1; 2.
 DR PROSITE; PS50068; LDLRA_2; 4.
 DR PROSITE; PS0240; TRYPSIN_DOM; 1.
 KW Hydrolase; Protease; Serine protease.
 SQ SEQUENCE 663 AA; 73914 MW; 46B2A56C657C7739 CRC64;

Query Match 42.1%; Score 1969; DB 2; Length 663;
 Best Local Similarity 42.8%; Pred. No. 9.1e-127;
 Matches 363; Conservative 122; Mismatches 171; Indels 192; Gaps 6;
 QY 14 KDFGAGLKYNSRHEKVGLEEGVEFLPVNNVKKVKGPGRWVLAALVGLLVLGIG 73
 DB 2 KDSMMKYNRRPQSMNGFEGVEFLPAANTKKVKGAPKKGLAIFGVVIGALLSLITIG 61

```

QY 74 FLVWHLYQYDRVYQKVFNGYMRITNENFVDAYENSTEFVSLASKVKDALKLYSGVPP 133
DB 62 LLVWHFAYNAPVQKLYTGLYRIANTOFVEAVENSTREFADLSVKVISTLTTLNGEKD 121
QY 134 LQPYHKESAVTAPSEGS----VIAYWSEFSIPQHLVVEAEARVMAEVRVMLPPRARSILKS 190
DB 122 IAPYLOQCSISAFSEGSNNVVGYYWSEFSVPAFREAEKAISE----LKLPTVNLQR 178
QY 191 FVTVSWAPPTDSTQVORTQDNCSFGLHARGVELMRTFTTRFPDPSYPAHARCOWALRG 250
DB 179 FAVDSLVAVPTDQIARNFNKSCAFFLHSSAGVMTKFSPPGPTTPPPNARCILWTLRA 238
QY 251 DADSVLSLTFRSGFDLASCDERGSIDLVTYNTLSPMEPHALVOLCTGYPSPYNLTTHSSON 310
DB 239 DAGOMIRLAKFKYKMEKKCANAGDFVMVYDSLSPIEPRAQIRLCIGIYPPSYNLTFSSN 298
QY 311 VLLITLITNTERHGFETATFQOLPRMSSCGRLKKAQGTFNPSYPGHPNIDCTWNI 370
DB 299 VMLVTLVTDNVGKFFGLAEFSQFPKTSLCGGYIRDASGVFTSPYFPFGHPKIECIWDI 358
QY 371 EVNNOHKVSFKFFVLLPFGVAGTCTPKDYVEINGEKYCGERSOFVNTSNKITVRFH 430
DB 359 QVPDNKPKLRNFMFLAEPGVPTVKCTKDFEINGQYCKGERKFFVNNSSNKSMSRFEV 418
QY 431 SDQSYDTGFLAELYSYSSDDPCPGQFTCRIGRCIRKELRCRDGADCTDHDDELNCSDA 490
DB 419 SDQSYDTGFLAELYSYSSDDPCPGQFTCRIGRCIRKELRCRDGADCTDHDDELNCSDA 478
QY 491 GHQFTCKN-KFCKPLFWVCDSDVNDGDSDEGSCPAQTFRCNSGKCLSKSQCCNGKDD 549
DB 479 -LQFRCTNSKLCPSYFVCDGVDGSDDELACQCPNNTYKCGNGKCIPEQKCDRTDN 537
QY 550 CGDGSDEASCPKVVVTKHYRCINGLCLSKNGPCDCKGDCSDGSDSDE--KDCDCGLR 607
DB 538 CGDGSDEASCPKVVVTKHYRCINGLCLSKNGPCDCKGDCSDGSDSDEI-SAKNCNGR 597
QY 608 SFTROARVVGTDADGEWPMQVSLHALQGHQICGASLSPNWLVSAAHCYIDDRGPRYS 667
DB 598 PTKKSRIVGVNA----- 611
QY 668 DPTOWTAFGLHIDQORSAPGVQERRKRIISHPFNDFTFDYDIALLELEKPAEYSSMV 727
DB 612 ----- 611
QY 728 RPICLPDASHVFPACKATWVTGWGHTQYGGTGALILQGEIRVINQITCENLLPQOITPR 787
DB 612 ----- 611
QY 788 MMCVFLSGGVDSQCGDGGPLSSVEADGRIFQAGVWSGDCGCAORNKPGVYVTRLPFRD 847
DB 612 -----DSGGPLSSVELNKNVYLAGIVSWGEGCARRNKPQVYVTRVAMWRD 655
QY 848 WIKENTGV 855
DB 656 WIRDKTGL 663

```

RESULT 9

```

ID Q6PF94 PRELIMINARY; PRT; 799 AA.
AC Q6PF94;
DT 05-JUL-2004 (TRENBLrel. 27, Created)
DT 05-JUL-2004 (TRENBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TRENBLrel. 27, Last annotation update)
DE Tmpres66 protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Liver;

```

```

RX MEDLINE=22389257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Fellings E.A., Grouse L.H., Derge J.G., Schuler G.D.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grinwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Kryzhanovskii M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Liver;
RA Strausberg R.;
RL Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Belongs to peptidase family S1.
DR EMBL; BC057674; AAH57674.1; -.
DR HSSP; P20160; 1AE5.
DR GO; GO:0005886; C:plasma membrane; IDA.
DR InterPro; IPR000859; CUB.
DR InterPro; IPR002172; LDL receptor A.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR InterPro; IPR009003; Pept Ser Cys.
DR Pfam; PF00057; Ldl recept a; 1.
DR Pfam; PF00089; Trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00261; LDLRECEPTOR.
DR SMART; SM00192; LDLa; 3.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS01180; CUB; 1.
DR PROSITE; PS01209; LDLRA_1; 1.
DR PROSITE; PS50068; LDLRA_2; 3.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolyase; Protease; Serine protease.
SQ SEQUENCE 799 AA; 89557 MW; 16315A6464D5288 CRC64;

```

```

Query Match 24.0%; Score 1124.5; DB 2; Length 799;
Best Local Similarity 31.6%; Pred. No. 1.2e-68;
Matches 266; Conservative 142; Mismatches 345; Indels 90; Gaps 25;
QY 33 EEGVE----FLPVNNVKKVKGHPGWVLAALVIGLLLVLLGIGFGLVWH-LOYR-DVRV 86
DB 20 EEAAEPGEKFKPKNTKR-----KNRDYRPTLLVLAALVSGVMLTFLGKAEVTV 74
QY 87 QKVFNGYMRITNENFVDAYENSTEFVSLASKVKDALKLYSGVPPFLGPHKESAVTAF 146
DB 75 SQVYSGSLRVLNRHFSQDLGRRESIAFRSESAQAQMLQELVASTR-LGVYVNSSSVVSF 133
QY 147 SEGSVIAYWSEFSIPQHLVVEAEARVMAEVRV--MLPPRARSILKSVTVSVVAFPTDSK 204
DB 134 GEGPLTCFFWFILDIPEY-----QRLTSPSEVRELLVDELLSNSSTLASVKTGEYVDPE 188
QY 205 --TVQRTQDNCSFGLHARGVELMRTFTTPG--FP-DSPPFAHARCOWALRGDADSVLSLT 259
DB 189 GLVILEASVNDIVVLNSTGLCYTRYSVNPGQVPLKGPDOQTTSCLWHLOQPEDLMIKVR 248
QY 260 FRSFDLASCDERGSIDLVTYNTLSPMEPHALVOLCTGYPSPYNLTTHSSQNVLITLN 319
DB 249 L-EWTRVDCRDR----VAMYDAAGPLEKRLITSYVGCGRQEPVMEVLASGSVMAVWVKKG 303

```


FT	ACT SITE	762	762	Charge relay system (by similarity).
FT	CARBOHYD	136	136	N-linked (GlcNAc. . .) (Potential).
FT	CARBOHYD	184	184	N-linked (GlcNAc. . .) (Potential).
FT	CARBOHYD	216	216	N-linked (GlcNAc. . .) (Potential).
FT	CARBOHYD	338	338	N-linked (GlcNAc. . .) (Potential).
FT	CARBOHYD	433	433	N-linked (GlcNAc. . .) (Potential).
FT	CARBOHYD	453	453	N-linked (GlcNAc. . .) (Potential).
FT	CARBOHYD	518	518	N-linked (GlcNAc. . .) (Potential).
FT	VARSPLIC	409	461	LCGRILQYAEIRPVYTAGITINFISQISLTGVRVHY LGNOSDPCGE -> YHFLSSLMFLPPLPSLPSSTVTP SLEAQVNLRGARGASRGWCQACCP (in isoform 2).
FT	VARSPLIC	462	811	/FTid=VSP_008379. Missing (in isoform 2).
FT	CONFLICT	116	116	/FTid=VSP_008380. A -> V (in Ref. 4).
FT	SEQUENCE	811 AA;	89999 MW;	7EEF193F655DDE9D CRC64;
Query Match		24.0%;	Score 1124;	DB 1; Length 811;
Best Local Similarity		31.0%;	Pred. No. 1.4e-68;	
Matches		276;	Conservative 138;	Mismatches 327; Indels 148; Gaps 28;
Qy	11	GGPKDPCGAGLYNKRHEKVGLEBGEVFLPVNNVKKVEKHPGRMVVLAALVILGLLVL 70		
Db	20	GGQDGGDG-----EEAPEGFMKACE-----DSKKARGYLRLVFLVLL--ALLVLA 66		
Qy	71	GIGFLVWH-LQYR-DVRVQKVFNGYMRITNENFVDAYENSNSSTFVSLASKVKDKALKLY 128		
Db	67	SAGVLLWYFLGYKAEVWVSYSGSLVNLNHFQDLTRRESSAFRSETAKAQMKELI 126		
Qy	129	SGVFLGPKYKESAVTAFSEGSVIAYWSEFSIPQHLVEABRVMAEVRVMLPPRARSL 188		
Db	127	TSTR-LGTYTNSSSVSYSGEGLTCFWEFLQIPEH-----RRMLSPSEV.-V 171		
Qy	189	KSFVTVSWAPTDQKTVQRTQDNCSGFLHARGVELMRF-----TTPFPDPSY- 238		
Db	172	QALLVEELSTVNSAAVYRAE-----YVDPEGLVILEASVKDIAALNSTGLCYRYSYV 227		
Qy	239	-----PAH--ARCOMALRGDADSVLSLTFRSFDLASCDRSGDLTVVNTLSPME 286		
Db	228	GQGVRLKGPDLHASSCLWHLQPKMLKRL-ENTLAECRD-----LAWYDVAGPLE 282		
Qy	287	PHALVQL--CQTPPSYNLT-----FHSSQNVLLITLITNTERRHPGEATF 331		
Db	283	KRLITSVYGCRCQPVVEVLASGAIMAVWCKGLHSYVDPVLSV-----QPVV 331		
Qy	332	FQLRPMSSCGRLRKAQGTNSPYPGHYPNIDCTWNIENPHQVKVSKFYLLLEPG 391		
Db	332	FQACEVNLTLNRLDSQGLSTPTYPFSYSPQTHCSWHLTVPSLDYGLALWFDAYALRRQ 391		
Qy	392	VPAGTCPKDYVEINGEKVC-----ERSQFVVTNSNKITVRPHSDQSYDTDTGLAE 443		
Db	392	KYDLPTCQGWITQNRLLGLRILOPYAEIRPVVATAG---ITINFISQISLTGPGVRVH 448		
Qy	444	YLSYDSSDDPCPGQTCRTGRCIRKELRCGWDCTDHSDELNCSCDAGHOFCKNRFCKP 503		
Db	449	YGLYNQSDPCGPEGLCSV-----NGLCVP 472		
Qy	504	LFVWCVSDVNDGNSDQGSQCPAQTRFC--SNKCLSKSOQCNKDKDQGSDEASCPKV 562		
Db	473	---ACDGVKDCPNGLDERNCVCRA--TFQCKEDSTCISLPKVCQDQPCDLNGSDEEQCE- 527		
Qy	563	NVVTCTKHTYRLNGLCLSGNPECDGKEDCDGSLRSTRQARVVGTTAD 622		
Db	528	-GVPCGTFITQCDRSCVKPNPCQDGRPCDRGSDDEHCDGLQG---PSSRIYGVAVSS 584		
Qy	623	EGEWPQVSLHALQGHGICGASLISPNWLVSAAHYIDDRGFYSDDPTQWTAFLGLHDQS 682		
Db	585	EGEWPQVQASLQVRGR-HICGALLIADRWITAAHCFOED---SWASTVLTWTLGKWNQN 640		
Qy	683	QRSAPGVQVRRRLKRIISHPFNDFTFDYDIALLEKPAEYSNWRPICLPDASHVFPAG 742		
Db	641	SR-WPGEVSPKVSRLLLHPYHEEDSHDYDVALQLDHPVVRSAARVPCVCLPARSHPFEPG 699		

Qy	743	KAIWVTGHTQYGGTGALILQKEIRVINQTTTCENLLPOQITPRMVCVFLSGVDSQC 802
Db	700	LHCWITWGALRGEGPISNALQKVQVLIIPDLCEVYRVQVTRMLCAGYRKGGKDACQ 759
Qy	803	GDSGGPLSSVEADGRIPQAGVSWGDCGAQRNKGVVYTRLPFRDWIKE 851
Db	760	GDSGGPLVCKALSGRWFLAGLSWGLGCGRPYFGVYTRITGVLSWQQ 808

RESULT 12

Q6UXD8	PRELIMINARY;	PRT;	802 AA.
AC	Q6UXD8;		
DT	05-JUL-2004 (T-EMBLrel. 27, Created)		
DT	05-JUL-2004 (T-EMBLrel. 27, Last sequence update)		
DT	05-JUL-2004 (T-EMBLrel. 27, Last annotation update)		
DE	PVAE354.		
GN	ORFNames=UNQ354;		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=22887296; PubMed=12975309; DOI=10.1101/gr.1293003;		
RA	Clark H.P., Gurney A.L., Abaya E., Baker K., Baldwin D., Brush J.,		
RA	Chen J., Chow B., Chui C., Crowley C., Currell B., Deuel B., Dowd P.,		
RA	Eaton D., Foster J.S., Grimaldi C., Gu Q., Hase P.E., Heldens S.,		
RA	Huang A., Kim H.S., Klimowski L., Jin Y., Johnson S., Lee J.,		
RA	Lewis L., Liao D., Mark M., Robbie E., Sanchez C., Schoenfeld J.,		
RA	Seshagiri S., Simmons L., Singh J., Smith V., Stinson J., Vagts A.,		
RA	Vandlen R., Watanabe C., Wiewand D., Woods K., Xie M.H., Yansura D.,		
RA	Yi S., Yu G., Yuan J., Zhang M., Zhang Z., Goddard A., Wood W.I.,		
RA	Godowski P.;		
RT	"The secreted protein discovery initiative (SPDI), a large-scale		
RT	effort to identify novel human secreted and transmembrane proteins: a		
RT	bioinformatics assessment."		
RL	Genome Res. 13:2265-2270(2003).		
CC	-1- SIMILARITY: Belongs to peptidase family S1.		
DR	EMBL; AY358398; AAQ88764.1; -.		
DR	HSSP; P20160; IAE5.		
DR	GO; GO:0004263; F:chymotrypsin activity; IEA.		
DR	GO; GO:0008233; F:peptidase activity; IEA.		
DR	GO; GO:0004295; F:trypsin activity; IEA.		
DR	GO; GO:0006508; P:proteolysis and peptidolysis; IEA.		
DR	InterPro; IPR000859; CUB.		
DR	InterPro; IPR002172; LDL receptor A.		
DR	InterPro; IPR001254; Peptidase S1.		
DR	InterPro; IPR001314; Peptidase S1a.		
DR	InterPro; IPR009003; Pept_Ser_Cys.		
DR	Pfam; PF00431; CUB; 1.		
DR	Pfam; PF00057; Ldl_recept_a; 1.		
DR	Pfam; PF00089; Trypsin; 1.		
DR	PRINTS; PR00722; CHYMOTRYPSIN.		
DR	PRINTS; PR00261; LDLRECEPTOR.		
DR	SMART; SM00192; LDLA; 3.		
DR	SMART; SM00020; Tryp_SPC; 1.		
DR	PROSITE; PS01180; CUB; 1.		
DR	PROSITE; PS01209; LDLRA_1; 1.		
DR	PROSITE; PS0068; LDLRA_2; 3.		
DR	PROSITE; PS0240; TRYP SIN DOM; 1.		
DR	PROSITE; PS00134; TRYP SIN HIS; UNKNOWN_1.		
DR	PROSITE; PS00135; TRYP SIN_SER; 1.		
SW	Hydrolase; Protease; Serine protease.		
Q6	SEQUENCE 802 AA; 88845 MW; 8726C91B1E02E163 CRC64;		

Query Match	24.0%;	Score 1122;	DB 2; Length 802;
Best Local Similarity	31.0%;	Pred. No. 1.9e-68;	
Matches	276;	Conservative 137;	Mismatches 328; Indels 148; Gaps 28;
Qy	11	GGPKDPCGAGLYNKRHEKVGLEBGEVFLPVNNVKKVEKHPGRMVVLAALVILGLLVL 70	


```
QY 504 LFWVCDVNDGNSDEQSCPAQTFRC-SNGKCLSKSQCCNGKDCDGDGSDASCPKV 562
Db 464 ---ACDGVKDCPNGLDERNCVCRA-TFOCKEDSTCISLPKVCDDGQDCLNGSDERQCB- 518
QY 563 NVVTCIKHYRCLNGCLSKNPECCKEDCDGSDGDEKDCDGLRSFTQARVVGTDAD 622
Db 519 -GVPCTFTFQCBDRSCVKPNQPCDGRPDGSDGDEHCDGLQ--PSSRIVGVGVSS 575
QY 623 EGEWPQVSHALGQGHICGASLISPNMLVSAHCVIDDGRFPRYSDPTQWTAFLGLHDS 682
Db 576 EGEWPQASLOVGR-HICGGALADRWTAAHCFQED---SMATVLTWTVFLKVVQN 631
QY 683 QRSAGVQVRRRLRIISHFFNDFTDYDIALLELEKPAEYSSMVRPCLPDASHVFPAG 742
Db 632 SR-WPGEVSKYRLLHPHYEEDSHDYDVALQLDHPVRSAAVRPCLPARSHFFPFG 690
QY 743 KAIWVGHTQYGGTGA-----LILQGEIRVINQTCENLL 780
Db 691 LHCWITGMALREGALRAADAVAFYGRNQGSETCCCPISNALOKVDVQLIPQDLCSVY 750
QY 781 POQITPRMVCVFLSGGVDSGCGPLSSVEADGRIFQAGVSWGDCGCAQRNKPQVYT 840
Db 751 RQVTPMLCAGYRKKKDACQSGGGLPVCKALSGRHLFAGLVSWGLGCGRPYFGVYT 810
QY 841 RLPLFRDWIKE 851
Db 811 RITGVISWTOQ 821

RESULT 14
QYRTY8
ID QYRTY8 PRELIMINARY; PRT; 572 AA.
AC QYRTY8;
DT 01-MAR-2004 (TremBLrel. 26, Created)
DT 01-MAR-2004 (TremBLrel. 26, Last sequence update)
DT 25-OCT-2004 (TremBLrel. 28, Last annotation update)
DE Type II transmembrane serine protease 7 precursor (Hypothetical
protein FLJ16088).
GN Name=TMPRSS7;
OS Homo sapiens (Human);
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2272134; PubMed=12838346;
RA Puentes X.S., Sanchez J.M., Overall C.M., Lopez-Otin C.;
RT "Human and mouse proteases: a comparative genomic approach.";
RL Nat. Rev. Genet. 4:544-558(2003).
RN [2]
RP SEQUENCE FROM N.A.
RA Ninomiya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,
RA Puruya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K.,
RA Katsuta N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K.,
RA Kanehori K., Takahashi-Pujji A., Oshima A., Sugiyama A., Kawakami B.,
RA Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Belongs to peptidase family S1.
DR EMBL; BN000125; CAD67577.1; -
DR EMBL; AK131211; BAD18401.1; -
DR HSSP; P00760; 1EXZ.
DR MEROPS; S01.072; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0004263; F:chymotrypsin activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000859; CUB.
DR InterPro; IPR002172; IRL_receptor_A.
DR InterPro; IPR001254; Peptidase_S1.
```

```
DR InterPro; IPR001314; Peptidase_S1A.
DR InterPro; IPR009003; Pept_Ser_Cys.
DR Pfam; PF00431; CUB; 1.
DR Pfam; PF00057; Ldl_recept_a; 2.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00261; LDLRECEPTOR.
DR PROSITE; PS01180; CUB; 2.
DR PROSITE; PS01209; LDLRA_1; 1.
DR PROSITE; PS00068; LDLRA_2; 2.
DR PROSITE; PS02040; TRYPsin_DOM; 1.
DR PROSITE; PS00134; TRYPsin_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPsin_SER; 1.
KW Hydrolase; Protease; Serine protease; Signal; Transmembrane.
FT SIGNAL 1 23 Potential.
FT CHAIN 24 572 type II transmembrane serine protease 7.
SQ SEQUENCE 572 AA; 64041 MW; 0458CBEA996EA285 CRC64;

Query Match 21.6%; Score 1011.5; DB 2; Length 572;
Best Local Similarity 36.6%; Pred. No. 4.8e-61;
Matches 228; Conservative 97; Mismatches 221; Indels 77; Gaps 23;

QY 244 CQWALRGDADSVLSITFRSPDLASCDERGSIDLVTYNTLSMPHALVOLCGTYPSPVNL 303
Db 2 CFFKLVAVGVYLRSLIKSIQI-EADNCVTDLSLTIYDLSLLPIRSSILYRIC--EPTRTLM 58
QY 304 TETHSSQNVLLITLTNTERRHGPGFEATFQOLPRMSSCGGLRK---AQGFNSPYYPGH 359
Db 59 SFVSTNNMLVTPKSPHRLSGIRAYFEVPEOKCENTVLVKDITGEGKISSPYPSY 118
QY 360 YPNIDICTWNIENNOHVKSFKFF-YLLEPGVPAGTCKPKDYVEINGEKTCGE--RSQF 416
Db 119 YPPKCKCTWKFTQ-SLSTGLTALKFYNSITTKSMKG-CEHGWEINEHMYCGSYMDHQ 176
QY 417 VVTSNSNKITVRHSDQSYDTGTGLABYLSYDSDSDPCP-GQFTRGRCIRKELRCDGWA 475
Db 177 IFRVPSLVHILQCSSRLSKPLLAEGYSYNTISQPCFVGSFRCSGLCVPAQRCDGVN 236
QY 476 DCTDHSDELNCSDAGHQFTCKNKG--KPLFWVCDVNDGNSDQSCGSCPAQTFRC 532
Db 237 DCFDESDEL-----FCVSPQ-----ACTSSPR- 260
QY 533 SNGKCLSKSQCCNGKDCDGDGSDASCPKVNVTCTKHYRCLNGCLSKNPECCKED 592
Db 261 QHGPLI-----CDGFRDCENGRDEQNC--TQSPICNNRTFCGNDICFRQKXACDGTVD 313
QY 593 CSPGSDKDCDGLRSFTQARVVGTDADGEGWPQVSHALGQGHICGASLISPNMLV 652
Db 314 CPDGSDEEGCTCS-RSSALHRIIGGDTTLEGGWPQVSLHFVGSAY-CGASVISREWLL 371
QY 653 SAACHYIDDRGFYSDPTQWTAFLGLHDSQRS--APGVQERRLKRIISHFFNDFTDY 710
Db 372 SAACHP---HGNRLSDPTFWTAHLGMVQGNKAFVSP-----VRRIVVHEYNSTQFDY 422
QY 711 DIALLELE--KPAEYSSMVRPCLPDASHVFPAGKAIWTVGWGHT-QYGFAGLILQKE 767
Db 423 DIALQLSLAMPETLQLIQCIPPTGQVRSEKCKWVTGWRHREADNKGSLVLAQAE 482
QY 768 IRVINQTCENLLPQQTIPRMVCVFLSGGVDSGCGPLSSVEADGRIFQAGVSW 826
Db 483 VELIDQTLVSTY-GIITSRLCAGISGRKDACGSDGSGPLSCRRKSDGKWLITGVSW 541
QY 827 GDGCAQRNKPQVYTRPLPRDWI 849
Db 542 GHGGRPNFPQVYTRVSNFVPMI 564

RESULT 15
QYRTY8
ID QYRTY8 PRELIMINARY; PRT; 572 AA.
AC QYRTY8;
DT 01-MAR-2003 (TremBLrel. 23, Created)
DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)
```

01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
Mus musculus adult male corpora quadrigenina cDNA, RIKEN full-length
enriched library, clone:B230219123 product:weakly similar to BLOOD
COAGULATION FACTOR XI
Name=Timprase7;
Mus musculus (Mouse).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigenina;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
[2]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigenina;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
[3]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigenina;
RA The FANTOM Consortium;
RT "Analysis of the mouse transcriptome based on functional annotation of
60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
[4]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigenina;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu M., Sugahara Y., Shibata K., Itoh M.,
Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
[5]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigenina;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
Fujitake S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
[6]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigenina;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
Kato H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
Saito R., Saichoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
Tagawa A., Takahashi E., Takaku-Akahara S., Takeda Y., Tanaka T.,
Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Belongs to peptidase family S1.
DR EMBL: AK045663; BAC32448.1; --
DR HSP: P00760; 1EXZ.
DR MEROPS: S01.072; --

MGD; MGI:2686594; Tmpras7.
GO: GO:0004263; P:chymotrypsin activity; IEA.
GO: GO:0008233; P:peptidase activity; IEA.
GO: GO:0004235; P:trypsin activity; IEA.
GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
InterPro: IPR000859; CUB.
InterPro: IPR002172; LDL_receptor_A.
InterPro: IPR001254; Peptidase_S1.
InterPro: IPR001314; Peptidase_S1A.
InterPro: IPR009003; Pept_Ser_Cys.
Pfam: PF00431; CUB; 1.
Pfam: PF00057; Ldl_recept_a; 3.
Pfam: PF00089; Trypsin; 1.
PRINTS: PR00722; CHYMOTRYPSIN.
PRINTS: PR00261; LDLRECEPTOR.
SMART: SM00042; CUB; 1.
SMART: SM00192; LDLa; 3.
SMART: SM00020; Tryp_Spc; 1.
PROSITE: PS01180; CUB; 2.
PROSITE: PS01209; LDLRA_1; 1.
PROSITE: PS00068; LDLRA_2; 2.
PROSITE: PS02040; TRYPsin_DOM; 1.
PROSITE: PS00134; TRYPsin_HIS; UNKNOWN_1.
PROSITE: PS00135; TRYPsin_SER; 1.
KW Hydrolase; Protease; Serine protease.
SQ SEQUENCE 572 AA; 63757 MW; 21C0EC523B5F4301 CRC64;

Query Match 21.6%; Score 1011.5; DB 2; Length 572;
Best Local Similarity 36.5%; Pred. No. 4.8e-61;
Matches 226; Conservative 96; Mismatches 227; Indels 71; Gaps 21;

Qy 244 COMALRGDADSVLSLTFERSFDLASCDBRGSDLVVYNTLSPMEHALVOLCGTYPPSYNL 303
Db 2 CFPKLVAVGVILRLSIESIQL-EADNCITDSLTVDLSLLPIRSAILYRIC--EPTRTLM 58

Qy 304 TFHSSQNVLLITLTNTERRHPGFEATFPOLPMSSCGG-----RLRKAQTFNSPPYPG 358
Db 59 SFVSTNNMLVLKSPYVRRLAGIRAYFEVIPE-QKCESTILVKEINSFEKGISPPYPS 117

Qy 359 HYPNIDICTWNIEVPNNQHVKVPFVLEPGVPAGT'CPKDYVEINGEKTCGE--RSQF 416
Db 118 YPPKCKCTWTFQT-SLSTIGIALKFPYNSITKSAKGEHWIENHMYCGSYMDHET 176

Qy 417 VVTSNSNKITVRHSDQSYTDTGFLAEYLSYSDSDPCP-GQFTCTGRCITKELRCGWA 475
Db 177 IFRVPSPLVHIQCSRLSDKPLLVYGGYVNISSQCPAGSFRCSGLCVPAQRCQGVN 236

Qy 476 DCTDHSDELNCSDAGHQFTCKNFKFPLFWVCDNVNDCGNSDEQSCSQAQTFRCNSG 535
Db 237 DCFDESDELFCVT-----VXP-----ACNSSSFR-OHG 263

Qy 536 KCLSKSQCCNGKDDCGDGSDEASC'PKVNVVTCTKHYRCLNGLCLSKGNPCDCKDCSD 595
Db 264 PLV-----CDGFRDCEGDEQCNTR--SIPCTSRTEKCNNDICFRKNAQCGIVDCPD 316

Qy 596 GSDEKDCDGLRSFTROARVVGTTDADGEWPMQVSLHALGQGHICGASLSPNWLVSAA 655
Db 317 GSDEEGCGCS-RSSSLFRLIRVGGSDSQEGTWPMQVSLHFGVSAY-CGASVLSREWLLSAA 374

Qy 656 HCYIDDRGFRYSPTQWTAFGLHDQSORS--APGVOERLKRILSHIFFNDFDFFDIA 713
Db 375 HCF---HGNRLSDPTPTWHLGMYVQGNAKFISP-----VRRIVVHEYNQSTFFDIA 425

Qy 714 LLELE--KPAEYSMWVRPICLPDASHVFPAGKAIWVTGWGHT-OYGGTGALILKQGIHV 770
Db 426 LLQLSTAWPETLKLQLOPICIIPAGQVRSRGEKCNWTCNGRRHEADSGSPVLQOAEVEL 485

Qy 771 INQTTCENLPLQOITPRMCMVGLSGVDSQCGDGGPLS-SVEADGRIFQAGVVSNGDG 829
Db 486 IDQTVCVSTY-GIITSRLCAGVMSGKSDACKGDSGGPLSCRRKSDGKWLITGIVSMGHG 544

Qy 830 CAQRNKGVTYTRLPFLFDWI 849
Db 1 : | | | | | : | | | | |

Fri Sep 23 14:45:56 2005

us-09-421-213-2.rup

Page 17

Db 545 CGRNPFGVYTRYSSFVPMI 564

Search completed: September 23, 2005, 12:59:09
Job time : 184 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 23, 2005, 12:46:25 ; Search time 82 Seconds
(without alignments)
4032.684 Million cell updates/sec

Title: US-09-421-213-2

Perfect score: 4681

Sequence: 1 MGSDRARKGGGPKDFGAGL.....PGVYTRFLFRDWIKENTGV 855

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : A_Geneseq_16Dec04:*

- 1: geneseqp1980s:*
- 2: geneseqp1990s:*
- 3: geneseqp2000s:*
- 4: geneseqp2001s:*
- 5: geneseqp2002s:*
- 6: geneseqp2003as:*
- 7: geneseqp2003bs:*
- 8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4681	100.0	855	2	AAY06671 Tumour an
2	4681	100.0	855	4	AAB98500 Human TAD
3	4681	100.0	855	4	AAB06930 Human mem
4	4681	100.0	855	5	AAO22929 Type II t
5	4681	100.0	855	5	ADi16816 Human NOV
6	4681	100.0	855	5	ADi16882 Human NOV
7	4681	100.0	855	5	ADi16875 Human NOV
8	4681	100.0	855	6	ABP56619 Human mem
9	4681	100.0	855	6	AAO30146 Human mem
10	4681	100.0	855	6	AAE29820 Human mem
11	4681	100.0	855	6	AAE29791 Human mem
12	4681	100.0	855	6	ABP72376 Transmem
13	4681	100.0	855	7	ADBP97551 Human MTS
14	4681	100.0	855	7	ADi110371 Human cel
15	4681	100.0	855	8	ADG65326 Human MTS
16	4681	100.0	855	8	ADi28861 Human mat
17	4681	100.0	855	8	ADJ46895 Human tra
18	4676	99.9	855	3	AAB19552 Human mat
19	4676	99.9	855	4	AAAB35465 Human mem
20	4676	99.9	855	5	ADi16817 Human NOV
21	4676	99.9	855	5	ADi16883 Human NOV
22	4676	99.9	855	5	ADi16876 Human NOV
23	4676	99.9	855	7	ADN39867 Cancer/an
24	4676	99.9	855	8	ADN04754 Antipsori
25	4676	99.9	855	8	ADP23334 PRO polyp

ALIGNMENTS

RESULT 1

RAY06671
ID RAY06671 standard; protein; 855 AA.

XX AC RAY06671;

DT 09-NOV-1999 (first entry)

DE Tumour antigen derived gene-15 (TAGD-15) protein.

XX Tumour antigen derived gene-15; TAGD-15; serine protease; human;
KW breast cancer; ovary cancer; carcinoma; diagnosis.

XX Homo sapiens.

Key	Location/Qualifiers
FT Domain	1..54 /note= "cytoplasmic domain"
FT Domain	55..213 /note= "transmembrane domain"
FT Modified-site	109..111 /note= "Asn is N-glycosylated"
FT Region	214..447 /note= "CUB repeat"
FT Modified-site	302..304 /note= "Asn is N-glycosylated"
FT Region	453..602 /note= "ligand-binding repeat (class A motif)"
FT Region	481..483 /note= "conserved SDE motif"
FT Region	518..520 /note= "conserved SDE motif"
FT Region	554..556 /note= "conserved SDE motif"
FT Region	597..599 /note= "conserved SDE motif"
FT Cleavage-site	614..615 /note= "catalytic domain"
FT Domain	615..855

XX WO9942120-A1.

XX 26-AUG-1999.

XX 18-FEB-1999; 99WO-US003436.

XX 20-FEB-1998; 98US-00027337.

XX

PA (UYAR-) UNIV ARKANSAS
 XX O'brien TJ, Tanimoto H;
 XX WPI; 1999-527418/44.
 DR N-PSDB; AAX87815.
 XX A new extracellular serine protease for diagnosis of neoplastic disease.
 PT Claim 3; Fig 10; 7lpp; English.
 XX
 XX The present sequence represents a novel human extracellular serine
 CC protease, termed tumor antigen derived gene-15 protein (see AAX06671),
 CC that is overexpressed in breast and ovarian carcinomas. The TAGD-15 gene
 CC (see AAX87815) can be used as a diagnostic and therapeutic target in
 CC ovarian carcinoma and other carcinomas including breast, prostate, lung
 CC and colon. The ligand binding domains of TAGD-15 may be valuable in the
 CC uptake of specific molecules into tumor cells. The invention also
 CC provides: a vector that is capable of expressing DNA encoding TAGD-15
 CC protein; host cells selected from bacterial cells (especially Escherichia
 CC coli), mammalian cells, plant cells and insect cells; and a method of
 CC detecting expression of TAGD-15 protein using a hybridisation probe
 XX
 XX Sequence 855 AA;
 SQ
 Query Match 100.0%; Score 4681; DB 2; Length 855;
 Best Local Similarity 100.0%; Pred. No. 2.1e-307;
 Matches 855; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MGS DRARKGGGPKDFGAGLKYNRHEKVNGLGEGVEFLPVNNVKKVKGHGRWVWVLA 60
 Db 1 MGS DRARKGGGPKDFGAGLKYNRHEKVNGLGEGVEFLPVNNVKKVKGHGRWVWVLA 60
 QY 61 VLIGLLVLLGIGFLVHLYQYRDVRVQKVPNGVMRTNENFVDAYENSSTEFVSLASKV 120
 Db 61 VLIGLLVLLGIGFLVHLYQYRDVRVQKVPNGVMRTNENFVDAYENSSTEFVSLASKV 120
 QY 121 KDALKLLYSGVPLFGPYKHESAVTAFSEGSVIAYYWSEFSIPQHLVEEAERYMAEERVVM 180
 Db 121 KDALKLLYSGVPLFGPYKHESAVTAFSEGSVIAYYWSEFSIPQHLVEEAERYMAEERVVM 180
 QY 181 LPPRARSLKSFVTSVVAFTDSKTQRTQDSCSFLHARGVELMFTTGPFPSPYPA 240
 Db 181 LPPRARSLKSFVTSVVAFTDSKTQRTQDSCSFLHARGVELMFTTGPFPSPYPA 240
 QY 241 HARCOWALRGDADSVLSLTFRSDLASCDERSGLVTVNTLSPMEPHALVOLCGTYPPS 300
 Db 241 HARCOWALRGDADSVLSLTFRSDLASCDERSGLVTVNTLSPMEPHALVOLCGTYPPS 300
 QY 301 YNLTFHSSQNVLLITLITNTERRHPGFEATFFQLPRMSSCGRLRKAQGTFFNSPYYPGHY 360
 Db 301 YNLTFHSSQNVLLITLITNTERRHPGFEATFFQLPRMSSCGRLRKAQGTFFNSPYYPGHY 360
 QY 361 PNINDCTWNIENPNQHVKSFKFFVLLPFGVPAGTCKPDYVEINGEKYCGERSQFVTS 420
 Db 361 PNINDCTWNIENPNQHVKSFKFFVLLPFGVPAGTCKPDYVEINGEKYCGERSQFVTS 420
 QY 421 NSNKITVRPHSDQSYDTGTFLAEYLSYSSDDPCPGQFTCTGCTCKELRCDCWDADCTDH 480
 Db 421 NSNKITVRPHSDQSYDTGTFLAEYLSYSSDDPCPGQFTCTGCTCKELRCDCWDADCTDH 480
 QY 481 SDLNCSCDAGHQFTCKNFKCPFLFWVCDSDVNDGDNDSDEQSCSPAQTFRCNSGKCLSK 540
 Db 481 SDLNCSCDAGHQFTCKNFKCPFLFWVCDSDVNDGDNDSDEQSCSPAQTFRCNSGKCLSK 540
 QY 541 SQQCNKGKDCGDSDEASCPKVVNTCTKHTRCLNGLCLSGNPECCKGDCSGDSBK 600
 Db 541 SQQCNKGKDCGDSDEASCPKVVNTCTKHTRCLNGLCLSGNPECCKGDCSGDSBK 600
 QY 601 DCDCLGRSFTROARVVGGTADGGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
 Db 601 DCDCLGRSFTROARVVGGTADGGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYID 660

QY 661 DRGFYSDPTQWTAFLGLHDQSRSPAGVQERRLRKRIISHPPFNDFTFDYDIALLELEXP 720
 Db 661 DRGFYSDPTQWTAFLGLHDQSRSPAGVQERRLRKRIISHPPFNDFTFDYDIALLELEXP 720
 QY 721 AEYSMSVRPICLPDASHVFPAGKAIWVTGWGHTQVGGTGALILQKGEIRVINQTTCEILL 780
 Db 721 AEYSMSVRPICLPDASHVFPAGKAIWVTGWGHTQVGGTGALILQKGEIRVINQTTCEILL 780
 QY 781 PQQITPRMVCVGLSGGVDSCQDGGPLSSVEADGRIFOAGVSVSWGDCGAORNKPGVYT 840
 Db 781 PQQITPRMVCVGLSGGVDSCQDGGPLSSVEADGRIFOAGVSVSWGDCGAORNKPGVYT 840
 QY 841 RLPLFRDWIKENTGV 855
 Db 841 RLPLFRDWIKENTGV 855
 RESULT 2
 AAB98500
 ID AAB98500 standard; protein; 855 AA.
 XX AAB98500;
 XX
 DT 03-AUG-2001 (first entry)
 XX
 DE Human TAGD-15.
 KW Human; TAGD-15; cytostatic; vaccine; ovarian tumour; cancer;
 KW tumour antigen-derived gene 15; extracellular serine protease.
 OS Homo sapiens.
 XX WO200129056-A1.
 XX
 PD 26-APR-2001.
 XX
 PF 20-OCT-2000; 2000MO-US029095.
 XX
 PR 20-OCT-1999; 99US-00421213.
 XX
 PA (UYAR-) UNIV ARKANSAS.
 XX
 PI O'brien TJ, Tanimoto H;
 XX
 DR WPI; 2001-381031/40.
 XX
 DR N-PSDB; AAX23601.
 XX
 PT Novel extracellular serine protease, termed tumor antigen-derived gene 15
 PT protein overexpressed in carcinomas and DNA encoding it, for diagnosis,
 PT treatment, prevention of cancer, particularly breast, ovarian cancer.
 XX
 PS Claim 11; Fig 2; 130pp; English.
 XX
 CC The present sequence represents human tumour antigen-derived gene 15
 CC (TAGD-15) protein. TAGD-15 is an extracellular serine protease. It was
 CC found that TAGD-15 is over-expressed in ovarian tumours. TAGD-15 protein
 CC or its fragments of 9-20 residues that lack TAGD-15 protease activity are
 CC useful for vaccinating an individual against TAGD-15, having, suspected
 CC of having or at risk of getting cancer. Furthermore, the TAGD-15 gene can
 CC be used as a diagnostic or therapeutic target in cancer
 XX
 SQ Sequence 855 AA;
 Query Match 100.0%; Score 4681; DB 4; Length 855;
 Best Local Similarity 100.0%; Pred. No. 2.1e-307;
 Matches 855; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MGS DRARKGGGPKDFGAGLKYNRHEKVNGLGEGVEFLPVNNVKKVKGHGRWVWVLA 60
 Db 1 MGS DRARKGGGPKDFGAGLKYNRHEKVNGLGEGVEFLPVNNVKKVKGHGRWVWVLA 60
 QY 61 VLIGLLVLLGIGFLVHLYQYRDVRVQKVPNGVMRTNENFVDAYENSSTEFVSLASKV 120

```
Db 61 VLIGLLLVLLGIGFLVHWHLYQYRDVVRVQKVFNGYMRITNENFVDAYENSNSSTFVSLASKV 120
Qy 121 KDALKLLYSGVPFLGYPYHKESAVTAFSEGSVLIAYYWSFESIPOHLVVEAEARVMAERVVM 180
Db 121 KDALKLLYSGVPFLGYPYHKESAVTAFSEGSVLIAYYWSFESIPOHLVVEAEARVMAERVVM 180
Qy 181 LPPRARSLSKFVTVSVVAPPTDSKTVORTQDNSCSFGLHARGVELMRFTTGGFPDPSYPA 240
Db 181 LPPRARSLSKFVTVSVVAPPTDSKTVORTQDNSCSFGLHARGVELMRFTTGGFPDPSYPA 240
Qy 241 HARCQWALRGDADSVLSLTFRSFDLASCDERSGLTVVNTLSPMEPHALVOLCGTYPSP 300
Db 241 HARCQWALRGDADSVLSLTFRSFDLASCDERSGLTVVNTLSPMEPHALVOLCGTYPSP 300
Qy 301 YNLTFHSSQNVLLITLITNTERHPGFEATFFQLPRMSSCGGLRKAQGTNSPYYPGHY 360
Db 301 YNLTFHSSQNVLLITLITNTERHPGFEATFFQLPRMSSCGGLRKAQGTNSPYYPGHY 360
Qy 361 PPNIDCTWNIEVPNNQHVKSFKFYLLPEPGVPAGTCTPKDYVEINGEKYCGERSQFVVT 420
Db 361 PPNIDCTWNIEVPNNQHVKSFKFYLLPEPGVPAGTCTPKDYVEINGEKYCGERSQFVVT 420
Qy 421 NSNKITVRPHSDQSYTDGFLAEYLSYDSSDPCGQFTCRGRCIRKELRCGWDADCTDH 480
Db 421 NSNKITVRPHSDQSYTDGFLAEYLSYDSSDPCGQFTCRGRCIRKELRCGWDADCTDH 480
Qy 481 SDELNCSCDAGHQFTCKNFKCKPLFWVCDVNDGNSDEQSCSPAQTFRCNSGKCLSK 540
Db 481 SDELNCSCDAGHQFTCKNFKCKPLFWVCDVNDGNSDEQSCSPAQTFRCNSGKCLSK 540
Qy 541 SQQNGKDDCGDSDASCPKVVNTCTKHTYRCLNGLCLSKNPECDGKEDCDSDGSEK 600
Db 541 SQQNGKDDCGDSDASCPKVVNTCTKHTYRCLNGLCLSKNPECDGKEDCDSDGSEK 600
Qy 601 DCDCGLRFTQARVVGTDGEGWPQVSIHALGQGHICGASLISPNWLVSAAHCYID 660
Db 601 DCDCGLRFTQARVVGTDGEGWPQVSIHALGQGHICGASLISPNWLVSAAHCYID 660
Qy 661 DRGFRYSPTQWTAFLGLHDQSRSAPGVQERRLKRIISHFFNDFTPDYDIALLEK 720
Db 661 DRGFRYSPTQWTAFLGLHDQSRSAPGVQERRLKRIISHFFNDFTPDYDIALLEK 720
Qy 721 AEYSMSVRPICLEFDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTCENLL 780
Db 721 AEYSMSVRPICLEFDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTCENLL 780
Qy 781 PQOITPRMCMVGHLSGGVDSQGDGGLSSVEADGRIFQAGVYSWGDCGCAQRNKP 840
Db 781 PQOITPRMCMVGHLSGGVDSQGDGGLSSVEADGRIFQAGVYSWGDCGCAQRNKP 840
Qy 841 RLPLFRDWIKENTGV 855
Db 841 RLPLFRDWIKENTGV 855

RESULT 3
ID AA06930
XX AA06930 standard; protein; 855 AA.
AC AA06930;
XX
XX 16-Oct-2001 (first entry)
XX
XX Human membrane-type serine protease (MTSP) 1.
XX Human; transmembrane serine protease; membrane-type serine protease;
XX MTSP; protease domain; neoplastic disease; tumour; cancer; cytostatic;
XX lung carcinoma; colon adenocarcinoma; ovarian carcinoma; gene therapy;
XX matriptase.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
```

```
FT Domain 615..855
FT /label= Protease_domain
XX WO200157194-A2.
XX 09-AUG-2001.
XX
XX 02-FEB-2001; 2001WO-US003471.
XX
XX 03-FEB-2000; 2000US-0179982P.
XX 18-FEB-2000; 2000US-0183542P.
XX 22-JUN-2000; 2000US-0213124P.
XX 26-JUL-2000; 2000US-0220970P.
XX 08-SEP-2000; 2000US-00657986.
XX 22-SEP-2000; 2000US-0234840P.
XX (CORV-) CORVAS INT INC.
XX
XX Madison EL, Ong EO, Yeh J;
XX WPI; 2001-488877/53.
XX N-PSDB; AAD13113.
XX
XX Novel single chain polypeptide comprising protease domain of type-II
XX membrane-type serine protease or its catalytically active portion useful
XX for treating and preventing cancer and tumor.
XX
XX Claim 12; Page 195-197; 256pp; English.
XX
XX The invention relates to transmembrane serine proteases and their
XX corresponding nucleotides and the protease domain of a type-II membrane-
XX type serine protease (MTSP). MTSP is useful for identifying compounds
XX that modulate or inhibit its proteolytic activity and for formulating a
XX medicament for treating neoplastic disease. MTSP and its corresponding
XX nucleotides are useful in preventing or treating tumours or cancers such
XX as lung carcinoma, colon adenocarcinoma and ovarian carcinoma, in
XX diagnostics and in hybridisation assays. MTSP is useful as a diagnostic
XX marker for tumour development, growth and/or progression and as
XX immunogens to generate antibodies that specifically bind to it. MTSP DNA
XX is useful in a yeast two-hybrid system and in gene therapy. The present
XX sequence is human MTSP1 protein (also called matriptase)
XX
XX Sequence 855 AA;
XX
XX Query Match 100.0%; Score 4681; DB 4; Length 855;
XX Best Local Similarity 100.0%; Pred. No. 2.1e-307;
XX Matches 855; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX Qy 1 MGSDRARKGGGPKDFGAGLKYSRHEKVNGLBEGVEFLPVNNVKYKKGPGRWVLLAA 60
XX Db 1 MGSDRARKGGGPKDFGAGLKYSRHEKVNGLBEGVEFLPVNNVKYKKGPGRWVLLAA 60
XX
XX Qy 61 VLIGLLLVLLGIGFLVHWHLYQYRDVVRVQKVFNGYMRITNENFVDAYENSNSSTFVSLASKV 120
XX Db 61 VLIGLLLVLLGIGFLVHWHLYQYRDVVRVQKVFNGYMRITNENFVDAYENSNSSTFVSLASKV 120
XX
XX Qy 121 KDALKLLYSGVPFLGYPYHKESAVTAFSEGSVLIAYYWSFESIPOHLVVEAEARVMAERVVM 180
XX Db 121 KDALKLLYSGVPFLGYPYHKESAVTAFSEGSVLIAYYWSFESIPOHLVVEAEARVMAERVVM 180
XX
XX Qy 181 LPPRARSLSKFVTVSVVAPPTDSKTVORTQDNSCSFGLHARGVELMRFTTGGFPDPSYPA 240
XX Db 181 LPPRARSLSKFVTVSVVAPPTDSKTVORTQDNSCSFGLHARGVELMRFTTGGFPDPSYPA 240
XX
XX Qy 241 HARCQWALRGDADSVLSLTFRSFDLASCDERSGLTVVNTLSPMEPHALVOLCGTYPSP 300
XX Db 241 HARCQWALRGDADSVLSLTFRSFDLASCDERSGLTVVNTLSPMEPHALVOLCGTYPSP 300
XX
XX Qy 301 YNLTFHSSQNVLLITLITNTERHPGFEATFFQLPRMSSCGGLRKAQGTNSPYYPGHY 360
XX Db 301 YNLTFHSSQNVLLITLITNTERHPGFEATFFQLPRMSSCGGLRKAQGTNSPYYPGHY 360
XX
XX Qy 361 PPNIDCTWNIEVPNNQHVKSFKFYLLPEPGVPAGTCTPKDYVEINGEKYCGERSQFVVT 420
```

Db 361 PPNIDCTWIEVFNQHVKSFKFYLLPFGVAGTCKPKDYVEINGEKYCGERSQFVVT 420
 QY 421 NSNKITVRFSHQSYTDTGFLAEYLSYDSGSDPCPGQFTCTGRCIRKELRCGWDADCTDH 480
 Db 421 NSNKITVRFSHQSYTDTGFLAEYLSYDSGSDPCPGQFTCTGRCIRKELRCGWDADCTDH 480
 QY 481 SDELNCSADAGHQTCKNFKCKPLFWVCDVNDGSDGSCPAQTFRCNSGKCLSK 540
 Db 481 SDELNCSADAGHQTCKNFKCKPLFWVCDVNDGSDGSCPAQTFRCNSGKCLSK 540
 QY 541 SOQCNGKDCGSDGSDGSCPAQTFRCNSGKCLSK 600
 Db 541 SOQCNGKDCGSDGSDGSCPAQTFRCNSGKCLSK 600
 QY 601 DCDGLRSFTRQARVVGTTDADEGEWPQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
 Db 601 DCDGLRSFTRQARVVGTTDADEGEWPQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
 QY 661 DRGFYSDPTQWTAFLGLHDQSORSAAGVOERRLKRIISHPFNDFTFDYDIALLELEKP 720
 Db 661 DRGFYSDPTQWTAFLGLHDQSORSAAGVOERRLKRIISHPFNDFTFDYDIALLELEKP 720
 QY 721 AYSVMVRPCLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTCENLL 780
 Db 721 AYSVMVRPCLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTCENLL 780
 QY 781 PQOITPRMVCVGLSGGVDSGCGPLSSVEADGRIFQAGVSWGDCGCAORNKPGVYT 840
 Db 781 PQOITPRMVCVGLSGGVDSGCGPLSSVEADGRIFQAGVSWGDCGCAORNKPGVYT 840
 QY 841 RLPLFRDWIKENTGV 855
 Db 841 RLPLFRDWIKENTGV 855

RESULT 4

ID AAO22929 standard; protein; 855 AA.
 AC AAO22929;
 XT 12-DEC-2002 (first entry)
 DE Type II transmembrane serine protease 1 protein SEQ ID No 2.
 KW Cytostatic; type-II membrane-type serine protease 7; MTSP7; malignancy;
 KW neoplastic disease; pre-malignant lesion; tumour; zymogen form; cancer;
 KW malignant; enzyme.
 OS Homo sapiens.
 PN WO200272786-A2.
 PD 19-SEP-2002.
 PF 13-MAR-2002; 2002WO-US007903.
 PR 13-MAR-2001; 2001US-0275592P.
 PA (CORV-) CORVAS INT INQ.
 PI Madison EL, Ong EO;
 DR WPI: 2002-732827/79.
 DR N-PSDB; AAL53444.
 PT New transmembrane serine protease 7 (MTSP7) polypeptide for diagnosing
 PT neoplastic diseases, monitoring tumor progress or therapeutic
 PT effectiveness, or identifying MTSP7 modulators for treating tumors or
 PT cancers.
 PS Disclosure; Page 172-174; 184pp; English.

XX The invention relates to a purified single or two-chain polypeptide,
 CC which comprises the protease domain of a type-II membrane-type serine
 CC protease 7 (MTSP7) or its catalytically active portion. The polypeptide
 CC comprising MTSP7 is useful for detecting or diagnosing a neoplastic
 CC disease, a pre-malignant lesion, a malignancy or other pathological
 CC condition in a subject. This polypeptide is also useful for monitoring
 CC tumour (e.g. tumor of the breast, cervix, prostate, lung, ovary or
 CC colon) progress and/or therapeutic effectiveness. The inhibitor of the
 CC polypeptide or the inhibitor of the polypeptide's zymogen form is useful
 CC for treating or preventing a neoplastic disease, or tumour initiation,
 CC growth or progression, or a (pre-)malignant condition. The polypeptide or
 CC polynucleotide is also useful for identifying modulators of MTSP7, which
 CC may be used to treat cancers or tumours. This sequence represents a
 CC protein of the type-II membrane-type serine protease 1 relating to the
 CC invention
 XX
 SQ Sequence 855 AA;
 Query Match 100.0%; Score 4681; DB 5; Length 855;
 Best Local Similarity 100.0%; Pred. No. 2.1e-307;
 Matches 855; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MGSDRARKGGGPKDFGAGLKYSRHEKVNGLSEGVFLPVNNYKVKVKEHGRWVLA 60
 Db 1 MGSDRARKGGGPKDFGAGLKYSRHEKVNGLSEGVFLPVNNYKVKVKEHGRWVLA 60
 QY 61 VLI GLLVLLIGLIGFLVWHLOVRVQKVFNGYMRITNENFVDAYENSNSFEVSLASKV 120
 Db 61 VLI GLLVLLIGLIGFLVWHLOVRVQKVFNGYMRITNENFVDAYENSNSFEVSLASKV 120
 QY 121 KDALKLYSGVPFLGYPYHKS AVTAFSEGSVIAYWSEFSIPQHLVEFAERVMABERVVM 180
 Db 121 KDALKLYSGVPFLGYPYHKS AVTAFSEGSVIAYWSEFSIPQHLVEFAERVMABERVVM 180
 QY 181 LPPRARSLSKSFVTVSVAFPTDSKTQVORTDONS CSFGLHARGVELMRFTTPEGDPSPYPA 240
 Db 181 LPPRARSLSKSFVTVSVAFPTDSKTQVORTDONS CSFGLHARGVELMRFTTPEGDPSPYPA 240
 QY 241 HARCQWALRGDADSVLSITFRSFDLASCDSRGSGLVTVYNTLSMPHEHALVQLCGTYP 300
 Db 241 HARCQWALRGDADSVLSITFRSFDLASCDSRGSGLVTVYNTLSMPHEHALVQLCGTYP 300
 QY 301 YNLTFHSSQNVLLITLTNTERRHPGPEATFPQLPRMSSCGGRLRKAQGTNSPYPGHY 360
 Db 301 YNLTFHSSQNVLLITLTNTERRHPGPEATFPQLPRMSSCGGRLRKAQGTNSPYPGHY 360
 QY 361 PPNIDCTWIEVFNQHVKSFKFYLLPFGVAGTCKPKDYVEINGEKYCGERSQFVVT 420
 Db 361 PPNIDCTWIEVFNQHVKSFKFYLLPFGVAGTCKPKDYVEINGEKYCGERSQFVVT 420
 QY 421 NSNKITVRFSHQSYTDTGFLAEYLSYDSGSDPCPGQFTCTGRCIRKELRCGWDADCTDH 480
 Db 421 NSNKITVRFSHQSYTDTGFLAEYLSYDSGSDPCPGQFTCTGRCIRKELRCGWDADCTDH 480
 QY 481 SDELNCSADAGHQTCKNFKCKPLFWVCDVNDGSDGSCPAQTFRCNSGKCLSK 540
 Db 481 SDELNCSADAGHQTCKNFKCKPLFWVCDVNDGSDGSCPAQTFRCNSGKCLSK 540
 QY 541 SOQCNGKDCGSDGSDGSCPAQTFRCNSGKCLSK 600
 Db 541 SOQCNGKDCGSDGSDGSCPAQTFRCNSGKCLSK 600
 QY 601 DCDGLRSFTRQARVVGTTDADEGEWPQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
 Db 601 DCDGLRSFTRQARVVGTTDADEGEWPQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
 QY 661 DRGFYSDPTQWTAFLGLHDQSORSAAGVOERRLKRIISHPFNDFTFDYDIALLELEKP 720
 Db 661 DRGFYSDPTQWTAFLGLHDQSORSAAGVOERRLKRIISHPFNDFTFDYDIALLELEKP 720
 QY 721 AYSVMVRPCLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGEIRVINQTCENLL 780

QY	121	KDALKLLYSGVPLGPIYHKSASVAFSEGSVIAYYWSSEFSIQHLVEEAERVMABERVVM	180
Db	121	KDALKLLYSGVPLGPIYHKSASVAFSEGSVIAYYWSSEFSIQHLVEEAERVMABERVVM	180
QY	181	LPPRARSLKSFVTSVVAFTDSTKVQRTQDSCSFGHARGVELMRFTTPGFPDPSYPYA	240
Db	181	LPPRARSLKSFVTSVVAFTDSTKVQRTQDSCSFGHARGVELMRFTTPGFPDPSYPYA	240
QY	241	HARQWALRGDADSVLSLFRSFDLASCDERSGLVTVNTLSMPHEPALVOLCGTYPPS	300
Db	241	HARQWALRGDADSVLSLFRSFDLASCDERSGLVTVNTLSMPHEPALVOLCGTYPPS	300
QY	301	YNLTFHSSQVLLITLITNTERRHGFCAETFFQLPRMSSCGGLRKAQGTFSPIYPGHY	360
Db	301	YNLTFHSSQVLLITLITNTERRHGFCAETFFQLPRMSSCGGLRKAQGTFSPIYPGHY	360
QY	361	PNIDCTWNIENPNQHVKSFKFFYLLPFGVPAGTCKPKDYVEINGEKYCGRSQFVVT	420
Db	361	PNIDCTWNIENPNQHVKSFKFFYLLPFGVPAGTCKPKDYVEINGEKYCGRSQFVVT	420
QY	421	NSNKITVRPHSDQSYTDTGFLAEYLSYSDSDPCGQFTCTGRCIRKELRCGWDCTDH	480
Db	421	NSNKITVRPHSDQSYTDTGFLAEYLSYSDSDPCGQFTCTGRCIRKELRCGWDCTDH	480
QY	481	SDELNCSCDAGHFTCKNFKCKPLFWVCDVNDGDNDSDEQSCSCPAQTFRCNSGKCLSK	540
Db	481	SDELNCSCDAGHFTCKNFKCKPLFWVCDVNDGDNDSDEQSCSCPAQTFRCNSGKCLSK	540
QY	541	SOQCKGKDCGDSDEASCPKVVNTCTKHTYRCLNGLCLSKGNPECDGKEDCSGDSBK	600
Db	541	SOQCKGKDCGDSDEASCPKVVNTCTKHTYRCLNGLCLSKGNPECDGKEDCSGDSBK	600
QY	601	DCDCGLRSFTRQARVVGTTDAGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYID	660
Db	601	DCDCGLRSFTRQARVVGTTDAGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYID	660
QY	661	DRGFRYSPTQWTAFLGLHDQSQRSAPGVQERLKRILISHPPFNDFTFDYDIALLEBKP	720
Db	661	DRGFRYSPTQWTAFLGLHDQSQRSAPGVQERLKRILISHPPFNDFTFDYDIALLEBKP	720
QY	721	BEYSSMVRPICLPDASHVFPAGKATWVTGWGHTQGGTGALILQGEIRVINQTCENLL	780
Db	721	BEYSSMVRPICLPDASHVFPAGKATWVTGWGHTQGGTGALILQGEIRVINQTCENLL	780
QY	781	POQITPRMCMVGLSGVDSCGDSGGLPSSVEADGRIFQAGVSWGDCQARNKPGVYT	840
Db	781	POQITPRMCMVGLSGVDSCGDSGGLPSSVEADGRIFQAGVSWGDCQARNKPGVYT	840
QY	841	RLPLFRDWIKENTGV 855	
Db	841	RLPLFRDWIKENTGV 855	

RESULT 6

AD116882
ID AD116882 standard; protein; 855 AA.

XX
AC AD116882;

DT 15-APR-2004 (first entry)

DE Human NOVX protein homologue SeqID 418.

KW human; NOVX; cardiomyopathy; atherosclerosis; cancer; diabetes;
KW inflammation; autoimmune disorder; allergy; blood disorder;
KW acquired immunodeficiency syndrome; AIDS; obesity; asthma;
KW immunoglobulin (Ig)A nephropathy; cirrhosis; arthritis;
KW Alzheimer's disease; infection; str.

OS Homo sapiens.

XX PN WO200268649-A2.

XX	06-SEP-2002.
PD	
XX	
PF	31-JAN-2002; 2002WO-US002785.
XX	
PR	31-JAN-2001; 2001US-0265395P.
PR	31-JAN-2001; 2001US-0265412P.
PR	31-JAN-2001; 2001US-0265514P.
PR	31-JAN-2001; 2001US-0265517P.
PR	02-FEB-2001; 2001US-0266406P.
PR	05-FEB-2001; 2001US-0266767P.
PR	07-FEB-2001; 2001US-0266975P.
PR	07-FEB-2001; 2001US-0267057P.
PR	08-FEB-2001; 2001US-0267459P.
PR	09-FEB-2001; 2001US-0267823P.
PR	15-FEB-2001; 2001US-0268974P.
PR	26-FEB-2001; 2001US-0271664P.
PR	27-FEB-2001; 2001US-0271839P.
PR	27-FEB-2001; 2001US-0271855P.
PR	02-MAR-2001; 2001US-0272788P.
PR	02-MAR-2001; 2001US-0273046P.
PR	14-MAR-2001; 2001US-0275925P.
PR	14-MAR-2001; 2001US-0275947P.
PR	14-MAR-2001; 2001US-0275950P.
PR	14-MAR-2001; 2001US-0275989P.
PR	15-MAR-2001; 2001US-0276448P.
PR	15-MAR-2001; 2001US-0276450P.
PR	16-MAR-2001; 2001US-0276397P.
PR	16-MAR-2001; 2001US-0276768P.
PR	20-MAR-2001; 2001US-0278652P.
PR	26-MAR-2001; 2001US-0278775P.
PR	26-MAR-2001; 2001US-0278778P.
PR	29-MAR-2001; 2001US-0279882P.
PR	29-MAR-2001; 2001US-0279884P.
PR	30-MAR-2001; 2001US-0280147P.
PR	11-APR-2001; 2001US-0282922P.
PR	11-APR-2001; 2001US-0283083P.
PR	20-APR-2001; 2001US-0285133P.
PR	23-APR-2001; 2001US-0285749P.
PR	03-MAY-2001; 2001US-0288327P.
PR	03-MAY-2001; 2001US-0288504P.
PR	29-MAY-2001; 2001US-0294047P.
PR	30-MAY-2001; 2001US-0294473P.
PR	08-JUN-2001; 2001US-0296964P.
PR	18-JUN-2001; 2001US-0298959P.
PR	19-JUN-2001; 2001US-0299324P.
PR	13-AUG-2001; 2001US-0312020P.
PR	16-AUG-2001; 2001US-0312889P.
PR	16-AUG-2001; 2001US-0312908P.
PR	21-AUG-2001; 2001US-0313390P.
PR	28-AUG-2001; 2001US-0315470P.
PR	31-AUG-2001; 2001US-0316447P.
PR	07-SEP-2001; 2001US-0318115P.
PR	07-SEP-2001; 2001US-0318118P.
PR	12-SEP-2001; 2001US-0318740P.
PR	19-SEP-2001; 2001US-0323379P.
PR	18-OCT-2001; 2001US-0330245P.
PR	18-OCT-2001; 2001US-0330308P.
PR	14-NOV-2001; 2001US-0332701P.

(CURA-) CURAGEN CORP.

Tchernev VT, Spytek KA, Zerhusen BD, Patturajan M, Shinkets RA;
Li L, Gangolli EA, Padigar M, Anderson DW, Rastelli L, Miller CE;
Gerlach VL, Taupier RJ, Gusev VY, Colman SD, Wolenc AR, Pena CEA;
Furtak K, Grosse WM, Alsobrook JP, Lepley DM, Rieger DK, Burgess CE;

WPI; 2002-706998/76.

New NOVX polypeptides and nucleic acids, useful for preventing or
treating NOVX-associated disorders, e.g. cancer, cardiomyopathy,
atherosclerosis, or diabetes, and in chromosome mapping, tissue typing or
pharmacogenomics.

XX PS Disclosure: SEQ ID NO 418; 1498pp; English.

XX CC This invention relates to a novel nucleic acids, and encoded polypeptides

CC thereof, which have properties related to the stimulation of biochemical

CC or physiological responses in a cell, tissue, organ or organism.

CC Specifically, it refers to the use of biologically active fragments for

CC diagnostic and prognostic assays and furthermore in the treatment of

CC diverse pathological conditions. The present invention describes novel

CC human and murine NOVX proteins, as well as methods to modulate their

CC expression using antisense oligos, ribozymes and peptide nucleic acids.

CC The NOVX polypeptides, polynucleotides and antibodies are useful in

CC treating or preventing NOVX-associated disorders, e.g. cardiomyopathy,

CC atherosclerosis, cancer and diabetes. Furthermore, they may be used in

CC treating or preventing diseases such as inflammation, autoimmune

CC disorders, allergies, blood disorders, acquired immunodeficiency syndrome

CC (AIDS), obesity, asthma, immunoglobulin (Ig)A nephropathy, cirrhosis,

CC arthritis, Alzheimer's disease, infections, stroke, muscular dystrophy

CC and epilepsy. Accordingly, these molecules have many activities including

CC cytoskeletal, cardiant, antiinflammatory, immunosuppressive, anti allergic,

CC haemostatic, anti-HIV, antidiabetic, antiarteriosclerotic, anorectic,

CC antiasthmatic, nephrotropic, antiarthritic, hepatotropic,

CC neuroprotective, nootropic, antibacterial, virucide, antiparasitic,

CC relaxant and anticonvulsant. In addition, they are useful in screening

CC assays to identify small molecules that modulate or inhibit, for example,

CC neurogenesis, wound healing and angiogenesis. The nucleic acids are also

CC used as in chromosome mapping, tissue typing, preventive medicine and

CC pharmacogenomics. This polypeptide is a homologue of a human NOVX protein

CC of the invention.

XX SQ Sequence 855 AA;

Query Match 100.0%; Score 4681; DB 5; Length 855;

Best Local Similarity 100.0%; Pred. No. 2.1e-307;

Matches 855; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGS DRARKGGGKDFGAGLKYNRSHKXVNGLEGEVFLPVNNVKKYKRGPGRWVLA 60

Db 1 MGS DRARKGGGKDFGAGLKYNRSHKXVNGLEGEVFLPVNNVKKYKRGPGRWVLA 60

Qy 61 VLIGLLVLIGLGLVHQLQYRDYRVQKVFNGYRNITNENFVDAYENSNSTEFVSLASKV 120

Db 61 VLIGLLVLIGLGLVHQLQYRDYRVQKVFNGYRNITNENFVDAYENSNSTEFVSLASKV 120

Qy 121 KDALKLYSGVPLFGPYHKGSAVTFSEGSAVIATYWSSEFSIPQHLVEAEKRWAEERVVM 180

Db 121 KDALKLYSGVPLFGPYHKGSAVTFSEGSAVIATYWSSEFSIPQHLVEAEKRWAEERVVM 180

Qy 181 LPPRARSLSKSFVMTSVVAFPTDSKTQVTDNSCSFGLHARGVELMFTTTPGFPDSPYPA 240

Db 181 LPPRARSLSKSFVMTSVVAFPTDSKTQVTDNSCSFGLHARGVELMFTTTPGFPDSPYPA 240

Qy 241 HARCQWALRGDASVLSITFRSFLASCDESGDLVTYNTLSPMEPHALVQLCGTYPPS 300

Db 241 HARCQWALRGDASVLSITFRSFLASCDESGDLVTYNTLSPMEPHALVQLCGTYPPS 300

Qy 301 YNLTFHSSQNVLLITLTNTERRHPGEATFFQPRMSSCGRLRKAQGTFFNSPYPGHY 360

Db 301 YNLTFHSSQNVLLITLTNTERRHPGEATFFQPRMSSCGRLRKAQGTFFNSPYPGHY 360

Qy 361 PPNIDCTWNIENRVNQHVKVSFKFFYLLEPGVPAGTCPKDYVEINGEKYCGERSQFVVT 420

Db 361 PPNIDCTWNIENRVNQHVKVSFKFFYLLEPGVPAGTCPKDYVEINGEKYCGERSQFVVT 420

Qy 421 NSNKITVRHSDGSYTDGTGLAEVLSYDSDPDPGQFTCTGRCIRKELRCDGWADCTDH 480

Db 421 NSNKITVRHSDGSYTDGTGLAEVLSYDSDPDPGQFTCTGRCIRKELRCDGWADCTDH 480

Qy 481 SDELNCSGAGHGTCKNKECKPLFWVCDSVNDGDNDSBQGCSCPAQTCRCNGKCLSK 540

Db 481 SDELNCSGAGHGTCKNKECKPLFWVCDSVNDGDNDSBQGCSCPAQTCRCNGKCLSK 540

Qy 541 SQQCNKGKDCGDSDEASCPKVNVTCTKHTYRCLNGLCLSKGNPECDGKEDCSGDSDEK 600

Db 541 SQQCNKGKDCGDSDEASCPKVNVTCTKHTYRCLNGLCLSKGNPECDGKEDCSGDSDEK 600

Qy 601 DCDGCLSFTRQARVVGTTDADEGEWPQVSLHALGQGHICGASLISPNWLVSAHACYID 660

Db 601 DCDGCLSFTRQARVVGTTDADEGEWPQVSLHALGQGHICGASLISPNWLVSAHACYID 660

Qy 661 DRGFRYSDPTQWTAFLGLHQDSQSAPGVORRLKRIISHPFFNDFTFDXDIALLELEKP 720

Db 661 DRGFRYSDPTQWTAFLGLHQDSQSAPGVORRLKRIISHPFFNDFTFDXDIALLELEKP 720

Qy 721 AYSYSSMVRPCLPDASHVFPAGKAIWVTGHTQYGGTGALILQKGEIRVNLQTTCCENLL 780

Db 721 AYSYSSMVRPCLPDASHVFPAGKAIWVTGHTQYGGTGALILQKGEIRVNLQTTCCENLL 780

Qy 781 PQQITPRMCMCVGLSGGVDSQCGSGPLSSVEADGRIFQAGVVSWGDCGCAQRNKPQYVT 840

Db 781 PQQITPRMCMCVGLSGGVDSQCGSGPLSSVEADGRIFQAGVVSWGDCGCAQRNKPQYVT 840

Qy 841 RLPLFRDWIKENTGV 855

Db 841 RLPLFRDWIKENTGV 855

RESULT 7

ADI16875

ID ADI16875 standard; protein; 855 AA.

XX AC ADI16875;

XX DT 15-APR-2004 (first entry)

XX DE Human NOVX protein homologue SeqID 411.

XX KW human; NOVX; cardiomyopathy; atherosclerosis; cancer; diabetes;

KW inflammation; autoimmune disorder; allergy; blood disorder;

KW acquired immunodeficiency syndrome; AIDS; obesity; asthma;

KW immunoglobulin (Ig)A nephropathy; cirrhosis; arthritis;

KW Alzheimer's disease; infection; str.

XX OS Homo sapiens.

XX PN WO200268649-A2.

XX PD 06-SEP-2002.

XX PF 31-JAN-2002; 2002WO-US002785.

XX 31-JAN-2001; 2001US-0265395P.

PR 31-JAN-2001; 2001US-0265412P.

PR 31-JAN-2001; 2001US-0265514P.

PR 31-JAN-2001; 2001US-0265517P.

PR 02-FEB-2001; 2001US-0266406P.

PR 05-FEB-2001; 2001US-0266767P.

PR 07-FEB-2001; 2001US-0266975P.

PR 07-FEB-2001; 2001US-0267057P.

PR 08-FEB-2001; 2001US-0267459P.

PR 09-FEB-2001; 2001US-0267823P.

PR 15-FEB-2001; 2001US-0268974P.

PR 26-FEB-2001; 2001US-0271664P.

PR 27-FEB-2001; 2001US-0271839P.

PR 27-FEB-2001; 2001US-0271855P.

PR 02-MAR-2001; 2001US-0272788P.

PR 02-MAR-2001; 2001US-0273046P.

PR 14-MAR-2001; 2001US-0275925P.

PR 14-MAR-2001; 2001US-0275947P.

PR 14-MAR-2001; 2001US-0275950P.

PR 14-MAR-2001; 2001US-0275989P.

PR 15-MAR-2001; 2001US-0276448P.

PR 15-MAR-2001; 2001US-0276450P.

PR 16-MAR-2001; 2001US-0276397P.

PR 16-MAR-2001; 2001US-0276768P.

PR 20-MAR-2001; 2001US-0278652P.

26-MAR-2001; 2001US-0278775P.
 26-MAR-2001; 2001US-0278778P.
 29-MAR-2001; 2001US-0279882P.
 29-MAR-2001; 2001US-0279884P.
 30-MAR-2001; 2001US-0280147P.
 11-APR-2001; 2001US-0282992P.
 11-APR-2001; 2001US-0283083P.
 20-APR-2001; 2001US-0285133P.
 23-APR-2001; 2001US-0285749P.
 03-MAY-2001; 2001US-0288327P.
 03-MAY-2001; 2001US-0288504P.
 29-MAY-2001; 2001US-0294047P.
 30-MAY-2001; 2001US-0294473P.
 08-JUN-2001; 2001US-0296964P.
 18-JUN-2001; 2001US-0298959P.
 19-JUN-2001; 2001US-0299324P.
 13-AUG-2001; 2001US-0312020P.
 16-AUG-2001; 2001US-0312889P.
 16-AUG-2001; 2001US-0312908P.
 21-AUG-2001; 2001US-0313390P.
 28-AUG-2001; 2001US-0315470P.
 31-AUG-2001; 2001US-0316447P.
 07-SEP-2001; 2001US-0318115P.
 07-SEP-2001; 2001US-0318118P.
 12-SEP-2001; 2001US-0318740P.
 19-SEP-2001; 2001US-0323379P.
 18-OCT-2001; 2001US-0330245P.
 18-OCT-2001; 2001US-0330308P.
 14-NOV-2001; 2001US-0332701P.
 (CURA-) CURAGEN CORP.
 Tchernev VT, Spytek KA, Zerhusen BD, Patturajan M, Shimkets RA, Li L, Gangolli EA, Padigaru M, Anderson DW, Rastelli L, Miller CE, Gerlach VL, Taupier RJ, Gusev VV, Colman SD, Wolenc AR, Pena CE, Furtak K, Grosse WM, Alsobrook JP, Lepley DM, Rieger DK, Burgess CE, WPI; 2002-706998/76.
 New NOVX polypeptides and nucleic acids, useful for preventing or treating NOVX-associated disorders, e.g. cancer, cardiomyopathy, atherosclerosis, or diabetes, and in chromosome mapping, tissue typing or pharmacogenomics.
 Disclosure; SEQ ID NO 411; 1499pp; English.
 This invention relates to a novel nucleic acids, and encoded polypeptides thereof, which have properties related to the stimulation of biochemical or physiological responses in a cell, tissue, organ or organism.
 Specifically, it refers to the use of biologically active fragments of diagnostic and prognostic assays and furthermore in the treatment of diverse pathological conditions. The present invention describes novel human and murine NOVX proteins, as well as methods to modulate their expression using antisense oligos, ribozymes and peptide nucleic acids. The NOVX polypeptides, polynucleotides and antibodies are useful in treating or preventing NOVX-associated disorders, e.g. cardiomyopathy, atherosclerosis, cancer and diabetes. Furthermore, they may be used in treating or preventing diseases such as inflammation, autoimmune disorders, allergies, blood disorders, acquired immunodeficiency syndrome (AIDS), obesity, asthma, immunoglobulin (Ig)A nephropathy, cirrhosis, arthritis, Alzheimer's disease, infections, stroke, muscular dystrophy and epilepsy. Accordingly, these molecules have many activities including cytostatic, cardiant, antiinflammatory, immunosuppressive, anti-allergic, haemostatic, anti-HIV, antidiabetic, antiarteriosclerotic, anorectic, antiasthmatic, nephrotropic, antiarthritic, hepatotropic, neuroprotective, nootropic, antibacterial, virucide, antiparasitic, relaxant and anticonvulsant. In addition, they are useful in screening assays to identify small molecules that modulate or inhibit, for example, neurogenesis, wound healing and angiogenesis. The nucleic acids are also used as in chromosome mapping, tissue typing, preventive medicine and pharmacogenomics. This polypeptide is a homologue of a human NOVX protein of the invention.

Qy	1	MGSDRARKGGGPKDFGAGLKYNSRHEKVNGLSEGVFLPVNNVKKVKGHGRWVLLAA	60
Db	1	MGSDRARKGGGPKDFGAGLKYNSRHEKVNGLSEGVFLPVNNVKKVKGHGRWVLLAA	60
Qy	61	VLIIGLLVLLIGLGVHQLYRDYRVQKVFNGYWRITNENFVDAYENSNSFEVSLASKV	120
Db	61	VLIIGLLVLLIGLGVHQLYRDYRVQKVFNGYWRITNENFVDAYENSNSFEVSLASKV	120
Qy	121	KDALKLYSGVFFLGPYHKSASVAFSEGSAVIAIYWSEFSIPQHLVEAEARVMAERVVM	180
Db	121	KDALKLYSGVFFLGPYHKSASVAFSEGSAVIAIYWSEFSIPQHLVEAEARVMAERVVM	180
Qy	181	LPPRARSLSKGFVTSVVAFPDTSKTQVORTQDNCSFGLHARGVELMRTTTFGFPDSPA	240
Db	181	LPPRARSLSKGFVTSVVAFPDTSKTQVORTQDNCSFGLHARGVELMRTTTFGFPDSPA	240
Qy	241	HARCOWALRGDADSVLSITFRSEDLASCDERGSGLVTVYNTLSMPHEHALVOLCGTYP	300
Db	241	HARCOWALRGDADSVLSITFRSEDLASCDERGSGLVTVYNTLSMPHEHALVOLCGTYP	300
Qy	301	YNLTFFHSSQNVLLITLTINTERRHPGFEATFFQPRMSSCGGRURKKAQGTENSPYPGHY	360
Db	301	YNLTFFHSSQNVLLITLTINTERRHPGFEATFFQPRMSSCGGRURKKAQGTENSPYPGHY	360
Qy	361	PNIDCTWNIEVNNQHVKSFKFYLLLEPGVPAGTCKDYVEINGEKYCGRSQFVVTS	420
Db	361	PNIDCTWNIEVNNQHVKSFKFYLLLEPGVPAGTCKDYVEINGEKYCGRSQFVVTS	420
Qy	421	NSNKITVRFHSDSYDTDTGFLAEVLSYDSDPCPGQTCRTGRCIRKELRCDGADCTDH	480
Db	421	NSNKITVRFHSDSYDTDTGFLAEVLSYDSDPCPGQTCRTGRCIRKELRCDGADCTDH	480
Qy	481	SDELNCSCDAGHOFCKNKECKPLFWYCDSDVNDGDNDSDEQSCPAQTFRCNSGKCLSK	540
Db	481	SDELNCSCDAGHOFCKNKECKPLFWYCDSDVNDGDNDSDEQSCPAQTFRCNSGKCLSK	540
Qy	541	SOQNGKDDCGSDSEASCPCNVVVTCTKTYRCLNGLCLSKGNPECDGKEDCDSDGSEK	600
Db	541	SOQNGKDDCGSDSEASCPCNVVVTCTKTYRCLNGLCLSKGNPECDGKEDCDSDGSEK	600
Qy	601	DCDGLSFTRQARVVGTTDADGEWPMQVSLHALGQGHICGASLISPNMLVSAHCYID	660
Db	601	DCDGLSFTRQARVVGTTDADGEWPMQVSLHALGQGHICGASLISPNMLVSAHCYID	660
Qy	661	DRGFYSDDPTQWTAFLGLHDQSRAPGVQERRLKRIISHPFFNFDFDYDIALLELEKP	720
Db	661	DRGFYSDDPTQWTAFLGLHDQSRAPGVQERRLKRIISHPFFNFDFDYDIALLELEKP	720
Qy	721	AEYSSWVRPCLPDASHVFPAGKAIWVTGHTGYGTGALILQKGBIRVINTTCENLL	780
Db	721	AEYSSWVRPCLPDASHVFPAGKAIWVTGHTGYGTGALILQKGBIRVINTTCENLL	780
Qy	781	PQOITPRMCMVGLSGVDSCQSDGGLSSVEADGRIFOAGVYVSWDGGCAQRNKPQVYT	840
Db	781	PQOITPRMCMVGLSGVDSCQSDGGLSSVEADGRIFOAGVYVSWDGGCAQRNKPQVYT	840
Qy	841	RLPLFRDWIKENTGV 855	
Db	841	RLPLFRDWIKENTGV 855	
RESULT 8			
ABP56619			
ID ABP56619 standard; protein; 855 AA.			
XX AC ABP56619;			
XX			

DT 24-MAR-2003 (first entry)
 XX Human membrane-type serine protease MTSPl protein SEQ ID NO:2.
 DE Human; membrane-type serine protease; enzyme; MTSPl0; cytosolic;
 KW type-II membrane-type serine protease; neoplastic disease; tumour; MTSPl;
 KW matriptase.
 XX
 OS Homo sapiens.
 OS
 PN WO200292841-A2.
 XX
 PD 21-NOV-2002.
 XX
 PF 14-MAY-2002; 2002WO-US015332.
 XX
 PR 14-MAY-2001; 2001US-0291001P.
 XX
 PA (CORV-) CORVAS INT INC.
 XX
 PI Madison EL, Yeh J;
 XX
 DR WPI; 2003-129309/12.
 XX
 DR N-PSDB; ABZ22450.
 XX
 PT New polypeptides comprising the protease domain of a type-II membrane-
 type serine protease (MTCPl0), or its muteins, useful for diagnosing
 PT neoplasms or malignancies, or for screening for MTCPl0 inhibitors for
 PT treating such diseases.
 XX
 PS Disclosure; Page 181-183; 198pp; English.
 XX
 CC The present invention describes a polypeptide comprising a purified
 CC single or two chain polypeptide, which comprises the protease domain of a
 CC type-II membrane-type serine protease (MTCPl0) or its catalytically
 CC active portion, or a mutein of it, where up to 50 % of the amino acids
 CC are replaced with another amino acid, and the resulting polypeptide is a
 CC single chain or two chain polypeptide that has a catalytic activity of at
 CC least 1-10 % of the unmutated polypeptide. MTCPl0 has cytosolic
 CC activity. The polypeptide containing the protease domain of the MTCPl0 is
 CC useful for detecting a neoplastic disease, and for diagnosing the
 CC presence of a pre-malignant lesion, a malignancy, or other pathological
 CC condition in a subject, or monitoring tumour (e.g. breast, cervix,
 CC prostate, lung, ovary or colon tumour) progression and/or therapeutic
 CC effectiveness. An inhibitor of the polypeptide containing the protease
 CC domain of MTCPl0 is useful for treating or preventing neoplastic disease
 CC in a mammal. An inhibitor of the activation cleavage of the zymogen form
 CC of the MTCPl0 polypeptide is useful for inhibiting tumour initiation,
 CC growth or progression, or treating (pre-)malignant conditions of the e.g.
 CC breast, cervix, prostate, lung, ovary or colon. The present sequence
 CC represents human MTSPl (also known as matriptase), which is used in an
 CC example from the present invention
 XX
 SQ Sequence 855 AA;

Query Match 100.0%; Score 4681; DB 6; Length 855;
 Best Local Similarity 100.0%; Pred. No. 2.1e-307;
 Matches 855; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGSDRARKGGGPKDFGAGLKYNSRHEKVGLEGEVFLPVNNVKKYKEHGPGRWVLA 60
 Db 1 MGSDRARKGGGPKDFGAGLKYNSRHEKVGLEGEVFLPVNNVKKYKEHGPGRWVLA 60

Qy 61 VLIGLLVLLVIGLGVVHQLQVRDVRQKVGNGYMRITNENFVDAYENSNSTFVSLASKV 120
 Db 61 VLIGLLVLLVIGLGVVHQLQVRDVRQKVGNGYMRITNENFVDAYENSNSTFVSLASKV 120

Qy 121 KDALKLLYSYGVPLFGPHKESAVTAFSEGSGVIAYTWSEFSPQHLVEEAERVMVAERVVM 180
 Db 121 KDALKLLYSYGVPLFGPHKESAVTAFSEGSGVIAYTWSEFSPQHLVEEAERVMVAERVVM 180

Qy 181 LPPRARSLSKSFVTSVVAFFPDTSKTVORTQDSCSFGHLHARGVELMRFTTGPFPDPSYPA 240
 Db 181 LPPRARSLSKSFVTSVVAFFPDTSKTVORTQDSCSFGHLHARGVELMRFTTGPFPDPSYPA 240

Db 181 LPPRARSLSKSFVTSVVAFFPDTSKTVORTQDSCSFGHLHARGVELMRFTTGPFPDPSYPA 240
 Qy 241 HARCQWALRGDADSVLSLTFRSPDLASCDERGSDDLVTYNTLSPMEPHALVOLCGTYPPS 300
 Db 241 HARCQWALRGDADSVLSLTFRSPDLASCDERGSDDLVTYNTLSPMEPHALVOLCGTYPPS 300
 Qy 301 YNLTFHSSQNVLLITLINTERRHPGFEATFQLPMSSCGGRLRKAQGTNSPYPGHY 360
 Db 301 YNLTFHSSQNVLLITLINTERRHPGFEATFQLPMSSCGGRLRKAQGTNSPYPGHY 360
 Qy 361 PPNIDCTWNIETVNNQHVKSFKFYLLEPGVPAGTCTPKDYVEINGEKYCGERSQFVVT 420
 Db 361 PPNIDCTWNIETVNNQHVKSFKFYLLEPGVPAGTCTPKDYVEINGEKYCGERSQFVVT 420
 Qy 421 NSNKITVRFHSDQSYTDTGFLAEYLSYDSSDPCQFTCTGRCIRKELRCDGWADCTDH 480
 Db 421 NSNKITVRFHSDQSYTDTGFLAEYLSYDSSDPCQFTCTGRCIRKELRCDGWADCTDH 480
 Qy 481 SDELNCSADAGHOFTCKNFKCKPLFWVCDVNDGDNDEQGCSCPAQTFRCNSGKCLSK 540
 Db 481 SDELNCSADAGHOFTCKNFKCKPLFWVCDVNDGDNDEQGCSCPAQTFRCNSGKCLSK 540
 Qy 541 SQQNGKDCDGDSDDEASCPKVNVTCTKTYRCLNGCLCLSKGNPECDGKDCSDGSEK 600
 Db 541 SQQNGKDCDGDSDDEASCPKVNVTCTKTYRCLNGCLCLSKGNPECDGKDCSDGSEK 600
 Qy 601 DDCGLRSFTRQARVVGTDADSGEWPQVSLHALGQGHICGASLISPNMLVSAHACYID 660
 Db 601 DDCGLRSFTRQARVVGTDADSGEWPQVSLHALGQGHICGASLISPNMLVSAHACYID 660
 Qy 661 DRGFYSDPTQWTAFLGLHDQSORSAFGVQERLKRILISHPFFNDFDIDALLELEKP 720
 Db 661 DRGFYSDPTQWTAFLGLHDQSORSAFGVQERLKRILISHPFFNDFDIDALLELEKP 720
 Qy 721 AEYSSMVRPCLPDASHVFPAGKAIWVTGHTQYGTGALILQKGEIRVNTQTCENLL 780
 Db 721 AEYSSMVRPCLPDASHVFPAGKAIWVTGHTQYGTGALILQKGEIRVNTQTCENLL 780
 Qy 781 PQQITPRMCMVGLSGGVDSQCGDSCGSLSSVEADGRIFOAGVVSWDGCGAQRNKPQVYT 840
 Db 781 PQQITPRMCMVGLSGGVDSQCGDSCGSLSSVEADGRIFOAGVVSWDGCGAQRNKPQVYT 840

RESULT 9
 AAO30146
 ID AAO30146 standard; protein; 855 AA.
 XX
 AC AAO30146;
 XX
 DT 03-SEP-2003 (first entry)
 XX Human membrane-type serine protease MTSPl protein.
 DE Serine protease 17; CVSP17; tumour; cancer; antisense therapy; prostate;
 KW breast; cervix; lung; ovary; colon; gene therapy; human; enzyme; MTSPl;
 KW membrane-type serine protease; matriptase.
 XX
 OS Homo sapiens.
 XX
 PN WO2003044179-A2.
 XX
 PD 30-MAY-2003.
 XX
 PF 20-NOV-2002; 2002WO-US037626.
 XX
 PR 20-NOV-2001; 2001US-0332015P.
 XX
 PA (CORV-) CORVAS INT INC.
 XX

Matches 855; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Qy	1 MGSDRARKGCGGPKDFGAGLKYNRHEKVNGLKEGVFLPVNNVKVKEHGPGRWVLA 60
Db	1 MGSDRARKGCGGPKDFGAGLKYNRHEKVNGLKEGVFLPVNNVKVKEHGPGRWVLA 60
Qy	61 VLIIGLLVLLGIGFLVHLYRDVVRQKVFNGYMRITNENFVDAYENSNSFEFVSLASKV 120
Db	61 VLIIGLLVLLGIGFLVHLYRDVVRQKVFNGYMRITNENFVDAYENSNSFEFVSLASKV 120
Qy	121 KDALKLLYSGVPLGPHYKESAVTAFSEGSVYAYWSEFSIPOHLVEEAERMAERVVM 180
Db	121 KDALKLLYSGVPLGPHYKESAVTAFSEGSVYAYWSEFSIPOHLVEEAERMAERVVM 180
Qy	181 LPPRARSLSKFVVTSVVAFPTDSKTQTDNSCSFGLHARGVELMRFTTGFDPSPYPA 240
Db	181 LPPRARSLSKFVVTSVVAFPTDSKTQTDNSCSFGLHARGVELMRFTTGFDPSPYPA 240
Qy	241 HARCOWALRGDADSVLSLTFRSFDLASCDSRSDLVTVYNTLSPMEPHALVOLCGTTPPS 300
Db	241 HARCOWALRGDADSVLSLTFRSFDLASCDSRSDLVTVYNTLSPMEPHALVOLCGTTPPS 300
Qy	301 YNLTFHSSQNVLLITLINTERRHPGFEATFQLPMSSCGRLRKAQGTNSPYPGHY 360
Db	301 YNLTFHSSQNVLLITLINTERRHPGFEATFQLPMSSCGRLRKAQGTNSPYPGHY 360
Qy	361 PPNDICTWNIENVNQHVKVSFKFFYLLEPGVPAGTCKDYVEINGEKYCGERSQFVVT 420
Db	361 PPNDICTWNIENVNQHVKVSFKFFYLLEPGVPAGTCKDYVEINGEKYCGERSQFVVT 420
Qy	421 NSNKITVRPHSDQSYTDGTGFLAEYLSYDSDPCPGQFTCRGRCIRKELCDGWADCTDH 480
Db	421 NSNKITVRPHSDQSYTDGTGFLAEYLSYDSDPCPGQFTCRGRCIRKELCDGWADCTDH 480
Qy	481 SDELNCSADAGHGFCKKNKFKPLFWCDVNDGDNDEGSCSPAOTFRCNSGKCLSK 540
Db	481 SDELNCSADAGHGFCKKNKFKPLFWCDVNDGDNDEGSCSPAOTFRCNSGKCLSK 540
Qy	541 SQQCNKGDDCGSDSEASCPKVVVYCTKHYRCLNGLCLSKNPECDKEDCDGSD 600
Db	541 SQQCNKGDDCGSDSEASCPKVVVYCTKHYRCLNGLCLSKNPECDKEDCDGSD 600
Qy	601 DCDCGLRSFTQARVVGTDADGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
Db	601 DCDCGLRSFTQARVVGTDADGEWPMQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
Qy	661 DRGFRYSPTQWTAFLGLHDQSORSAPGVQERRLKRIISHPPFNDFTFDYDIALLEK 720
Db	661 DRGFRYSPTQWTAFLGLHDQSORSAPGVQERRLKRIISHPPFNDFTFDYDIALLEK 720
Qy	721 AEYSSMVRPICLEDASHVFPAGKAIWVTGWGHTQYGGTGALILQGEIRVINQTTCE 780
Db	721 AEYSSMVRPICLEDASHVFPAGKAIWVTGWGHTQYGGTGALILQGEIRVINQTTCE 780
Qy	781 PQOITPRMVCVGLSGGVSDSCQSGGPLSSVEADGRIFQAGVSWGDCQARNKPGY 840
Db	781 PQOITPRMVCVGLSGGVSDSCQSGGPLSSVEADGRIFQAGVSWGDCQARNKPGY 840
Qy	841 RLPLFRDWIKENTGV 855
Db	841 RLPLFRDWIKENTGV 855

RESULT 11
AAE29791 standard; protein; 855 AA.
XX
AC AAE29791;
XX
DT 24-FEB-2003 (first entry)
XX
DE Human membrane-type serine protease, MTSP1.
XX

KW	Human; serine protease 14; CVSP14; cancer; malignancy; breast; colon; gene therapy; cytostatic; membrane-type serine protease; enzyme; MTSP1.
KW	Homo sapiens.
OS	WO200277263-A2.
FN	03-OCT-2002.
PD	20-MAR-2002; 2002WO-US009039.
XX	22-MAR-2001; 2001US-0278166P.
XX	(CORV-) CORVAS INT INC.
PA	Madison EL, Yeh J;
XX	WPI; 2003-018938/01.
PI	N-PSDB; AAD47180.
DR	New purified CVSP14 polypeptide and encoding nucleic acid molecule, useful for diagnosing, preventing and/or treating disorders, such as cancers and malignancies of the breast, cervix, prostate, lung, ovary or colon.
DR	Disclosure; Page 171-173; 185pp; English.
PS	The invention relates to transmembrane serine protease 14 (CVSP14), its nucleic acid sequence and the method based on them. The methods and compositions of the invention are useful for diagnosing, preventing or and/or treating conditions associated with the aberrant expression of activity of the CVSP14 polypeptide, such as cancers and malignancies of the breast, cervix, prostate, lung, ovary or colon. The methods are also useful for identifying compounds that will modulate the protease activity of CVSP14 polypeptide and monitoring tumour progression and/or therapeutic effectiveness. CVSP14 DNA used in gene therapy. The present sequence is human membrane-type serine protease, MTSP1
XX	Sequence 855 AA;
XX	Query Match 100.0%; Score 4681; DB 6; Length 855;
XX	Best Local Similarity 100.0%; Pred. No. 2.1e-307;
XX	Matches 855; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1 MGSDRARKGCGGPKDFGAGLKYNRHEKVNGLKEGVFLPVNNVKVKEHGPGRWVLA 60
Db	1 MGSDRARKGCGGPKDFGAGLKYNRHEKVNGLKEGVFLPVNNVKVKEHGPGRWVLA 60
Qy	61 VLIIGLLVLLGIGFLVHLYRDVVRQKVFNGYMRITNENFVDAYENSNSFEFVSLASKV 120
Db	61 VLIIGLLVLLGIGFLVHLYRDVVRQKVFNGYMRITNENFVDAYENSNSFEFVSLASKV 120
Qy	121 KDALKLLYSGVPLGPHYKESAVTAFSEGSVYAYWSEFSIPOHLVEEAERMAERVVM 180
Db	121 KDALKLLYSGVPLGPHYKESAVTAFSEGSVYAYWSEFSIPOHLVEEAERMAERVVM 180
Qy	181 LPPRARSLSKFVVTSVVAFPTDSKTQTDNSCSFGLHARGVELMRFTTGFDPSPYPA 240
Db	181 LPPRARSLSKFVVTSVVAFPTDSKTQTDNSCSFGLHARGVELMRFTTGFDPSPYPA 240
Qy	241 HARCOWALRGDADSVLSLTFRSFDLASCDSRSDLVTVYNTLSPMEPHALVOLCGTTPPS 300
Db	241 HARCOWALRGDADSVLSLTFRSFDLASCDSRSDLVTVYNTLSPMEPHALVOLCGTTPPS 300
Qy	301 YNLTFHSSQNVLLITLINTERRHPGFEATFQLPMSSCGRLRKAQGTNSPYPGHY 360
Db	301 YNLTFHSSQNVLLITLINTERRHPGFEATFQLPMSSCGRLRKAQGTNSPYPGHY 360
Qy	361 PPNDICTWNIENVNQHVKVSFKFFYLLEPGVPAGTCKDYVEINGEKYCGERSQFVVT 420
Db	361 PPNDICTWNIENVNQHVKVSFKFFYLLEPGVPAGTCKDYVEINGEKYCGERSQFVVT 420
Qy	421 NSNKITVRPHSDQSYTDGTGFLAEYLSYDSDPCPGQFTCRGRCIRKELCDGWADCTDH 480

Db 421 NSNKITVRHSDQSYDTDTGFLAEYLSYSDSDPCPGQFTCTGRCIRKELRCGWDCTDH 480
 Qy 481 SDELNCSCDAGHQTCTCKNFKCKPLFWVCDVNDGNSDEQSCSPAOTFRCSNGKCLSK 540
 Db 481 SDELNCSCDAGHQTCTCKNFKCKPLFWVCDVNDGNSDEQSCSPAOTFRCSNGKCLSK 540
 Qy 541 SQQCNKDCGDSDEASCPKVNVTCTKHTYRCLNGLCLSKGNPECDGKEDCSGDSBK 600
 Db 541 SQQCNKDCGDSDEASCPKVNVTCTKHTYRCLNGLCLSKGNPECDGKEDCSGDSBK 600
 Qy 601 DCDGLRSFTQARVVGTTDADGEGWPQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
 Db 601 DCDGLRSFTQARVVGTTDADGEGWPQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
 Qy 661 DRGFRYSDPTQWTAFLGLHDQSQRSAPGVQERLKRILSHPPFNDFTDYDIALLELEKP 720
 Db 661 DRGFRYSDPTQWTAFLGLHDQSQRSAPGVQERLKRILSHPPFNDFTDYDIALLELEKP 720
 Qy 721 AEYSSMVRPCLPDASHVPPAGKAIWVTGWGHTQYGGTALILQKGEIRVINQTCENLL 780
 Db 721 AEYSSMVRPCLPDASHVPPAGKAIWVTGWGHTQYGGTALILQKGEIRVINQTCENLL 780
 Qy 781 PQOITPRMCMVGLSGGVDSGCGDGLSSVEADGRIFQAGVWSGDCQQRNKPQYVT 840
 Db 781 PQOITPRMCMVGLSGGVDSGCGDGLSSVEADGRIFQAGVWSGDCQQRNKPQYVT 840
 Qy 841 RLPLFRDWIKENTGV 855
 Db 841 RLPLFRDWIKENTGV 855

RESULT 12

ABP72376
 ID ABP72376 standard; protein; 855 AA.
 XX AC ABP72376;
 XX DT 13-MAY-2003 (first entry)
 XX DE Transmembrane serine protease 1 (MTP1).
 XX KW Human; transmembrane serine protease 1; MTP1; MTP20; enzyme;
 KW cystostatic; dermatological; cardiant; vulnary; ophthalmological;
 KW gene therapy.
 XX OS Homo sapiens.
 XX PN WO2003004681-A2.
 XX PD 16-JAN-2003.
 XX PF 03-JUL-2002; 2002WO-US021208.
 XX PR 03-JUL-2001; 2001US-0302939P.
 XX PA (CORV-) CORVAS INT INC.
 XX PI Madison EL, Ong EO;
 XX DR WPI; 2003-239207/23.
 XX DR N-PSDB; ABZ58500.
 XX New type-II membrane-type serine protease 20 polypeptides, useful for
 PT preparing a medicament for diagnosing, treating or preventing cancer,
 PT dermatological disorders, aberrant wound repairs or crest syndromes.
 XX PS Disclosure; Page 198-200; 216pp; English.
 XX CC The present sequence is the protein sequence of human type II
 CC transmembrane serine protease 1 (MTP1). The invention relates to novel
 CC human type II transmembrane serine protein 20 (MTP20) (see ABP72374) and
 CC nucleic acids encoding it (see ABZ58499). Also claimed are methods of

CC inhibiting tumour initiation, growth or progression by inhibiting MTP20
 CC activity, and of treating or preventing a disease or disorder associated
 CC with undesired and/or uncontrolled angiogenesis or neovascularisation,
 CC especially undesired angiogenesis associated with solid neoplasms,
 CC vascular malformations and cardiovascular disorders, chronic inflammatory
 CC diseases, aberrant wound repairs, circulatory disorders, crest syndromes,
 CC dermatological disorders and ocular disorders using an inhibitor of MSP20
 XX SQ Sequence 855 AA;
 Query Match 100.0%; Score 4681; DB 6; Length 855;
 Best Local Similarity 100.0%; Pred. No. 2.1e-307; Indels 0; Gaps 0;
 Matches 855; Conservative 0; Mismatches 0;
 Qy 1 MGSDRARKGGGPKDFGAGLKYNRSHKVNGLSEGVFLPVNNVKKVEKKGPRWVVLAA 60
 Db 1 MGSDRARKGGGPKDFGAGLKYNRSHKVNGLSEGVFLPVNNVKKVEKKGPRWVVLAA 60
 Qy 61 VLIGLLVLVIGLGVHQLQYRDVQKVFNGYMRITNENFVDAYENSNSTEFVSLASKV 120
 Db 61 VLIGLLVLVIGLGVHQLQYRDVQKVFNGYMRITNENFVDAYENSNSTEFVSLASKV 120
 Qy 121 KDALKLYSGVPFLGYPYHKSAVTAFSEGSVIAIYWSSEFSIPOHLVEAEERVAERVVM 180
 Db 121 KDALKLYSGVPFLGYPYHKSAVTAFSEGSVIAIYWSSEFSIPOHLVEAEERVAERVVM 180
 Qy 181 LPPRARSLSKSPVTVSVVAFPTDSKTQVORTQDNSCSFGLHARGVELMRTTTPGFPDSPYPA 240
 Db 181 LPPRARSLSKSPVTVSVVAFPTDSKTQVORTQDNSCSFGLHARGVELMRTTTPGFPDSPYPA 240
 Qy 241 HARCQWALRGDADSVLSLTFRSPDLASCDERGSDLVTYNTLSPMEPHALVOLCGTYPSPS 300
 Db 241 HARCQWALRGDADSVLSLTFRSPDLASCDERGSDLVTYNTLSPMEPHALVOLCGTYPSPS 300
 Qy 301 YNLTFHSSQNVLLTLITNTERRHPGFEATFFQLPMSSCGGRKKAQGTNSPYPGHY 360
 Db 301 YNLTFHSSQNVLLTLITNTERRHPGFEATFFQLPMSSCGGRKKAQGTNSPYPGHY 360
 Qy 361 PPNIDCTWNIEVPNNQHVKSFKFYLLEPGVPAGTCKDYVEINGEKYQGERSOQVVT 420
 Db 361 PPNIDCTWNIEVPNNQHVKSFKFYLLEPGVPAGTCKDYVEINGEKYQGERSOQVVT 420
 Qy 421 NSNKITVRHSDQSYDTDTGFLAEYLSYSDSDPCPGQFTCTGRCIRKELRCGWDCTDH 480
 Db 421 NSNKITVRHSDQSYDTDTGFLAEYLSYSDSDPCPGQFTCTGRCIRKELRCGWDCTDH 480
 Qy 481 SDELNCSCDAGHQTCTCKNFKCKPLFWVCDVNDGNSDEQSCSPAOTFRCSNGKCLSK 540
 Db 481 SDELNCSCDAGHQTCTCKNFKCKPLFWVCDVNDGNSDEQSCSPAOTFRCSNGKCLSK 540
 Qy 541 SQQCNKDCGDSDEASCPKVNVTCTKHTYRCLNGLCLSKGNPECDGKEDCSGDSBK 600
 Db 541 SQQCNKDCGDSDEASCPKVNVTCTKHTYRCLNGLCLSKGNPECDGKEDCSGDSBK 600
 Qy 601 DCDGLRSFTQARVVGTTDADGEGWPQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
 Db 601 DCDGLRSFTQARVVGTTDADGEGWPQVSLHALGQGHICGASLISPNWLVSAAHCYID 660
 Qy 661 DRGFRYSDPTQWTAFLGLHDQSQRSAPGVQERLKRILSHPPFNDFTDYDIALLELEKP 720
 Db 661 DRGFRYSDPTQWTAFLGLHDQSQRSAPGVQERLKRILSHPPFNDFTDYDIALLELEKP 720
 Qy 721 AEYSSMVRPCLPDASHVPPAGKAIWVTGWGHTQYGGTALILQKGEIRVINQTCENLL 780
 Db 721 AEYSSMVRPCLPDASHVPPAGKAIWVTGWGHTQYGGTALILQKGEIRVINQTCENLL 780
 Qy 781 PQOITPRMCMVGLSGGVDSGCGDGLSSVEADGRIFQAGVWSGDCQQRNKPQYVT 840
 Db 781 PQOITPRMCMVGLSGGVDSGCGDGLSSVEADGRIFQAGVWSGDCQQRNKPQYVT 840
 Qy 841 RLPLFRDWIKENTGV 855
 Db 841 RLPLFRDWIKENTGV 855

RESULT 13
ADB97551
ID ADB97551 standard; protein; 855 AA.
XX
AC ADB97551;
XX
DT 04-DEC-2003 (first entry)
XX
DE Human MTSPL, SEQ ID NO:2.
XX
KW Human; type I transmembrane serine protease 1; MTSPL; matrilysin;
KW serine protease; protease cleavage activation; diagnostic marker;
KW neoplastic disorder; cancer; breast; cervix; prostate; lung; ovary;
KW cytotactic; gene therapy; drug screening; tumour progression; monitoring;
KW enzyme.
XX
OS Homo sapiens.
XX
PN WO2003031585-A2.
XX
PD 17-APR-2003.
XX
PF 08-OCT-2002; 2002WO-US032417.
XX
PR 09-OCT-2001; 2001US-0328530P.
XX
PA (CORV-) CORVAS INT INC.
XX
PI Madison EL, Yeh J;
XX
DR WPI; 2003-393442/37.
XX
DR N-PSDB; ADB97550.
XX
PT New purified single- or two-chain polypeptide, useful for diagnosing,
PT preventing or treating cancer (e.g. colon cancer), comprises a protease
PT domain of a type-I membrane-type serine protease 25 or its catalytically
PT active portion.
XX
PS Disclosure; Page 179-181; 97pp; English.
XX
CC The invention relates to human type I transmembrane serine protease 25
CC (MTSP25; ADB97551) and polypeptides derived from it (e.g., ADB97555). The
CC MTSP25 gene is located on chromosome 12. MTSP25 is a serine protease
CC which is activated by protease cleavage to yield a two-chain protease
CC comprising an A chain and a B chain linked by a disulphide bond. MTSP25
CC is expressed or is active in tumour cells, and can therefore be used as a
CC diagnostic marker for certain cancers. The invention also encompasses
CC nucleic acids encoding an MTSP25 polypeptide (ADB97564, ADB97571,
CC ADB97554); nucleic acid vectors and host cells comprising an MTSP25
CC polynucleotide; a MTSP25 knockout animal; and an antibody specific for
CC either the single chain (zymogen) or two-chain (activated) form of
CC MTSP25. MTSP25 polypeptides are useful in diagnosing, preventing or
CC treating neoplastic diseases, such as cancer of the breast, cervix,
CC prostate, lung, ovary or colon. MTSP25 polypeptides may also be used for
CC identifying compounds that modulate the protease activity of the
CC polypeptide and for monitoring tumour progression and/or therapeutic
CC effectiveness. The present sequence represents the related protein, MTSP1
CC (also known as matrilysin).
XX
SQ Sequence 855 AA;
Query Match 100.0%; Score 4681; DB 7; Length 855;
Best Local Similarity 100.0%; Pred. No. 2.1e-307; Indels 0; Gaps 0;
Matches 855; Conservative 0; Mismatches 0;
QY 1 MGSDRARGGGGKDFGAGLKYNSRHEKVNGLSEGVFLPNNVKYKHGPGRWVLLAA 60
DB 1 MGSDRARGGGGKDFGAGLKYNSRHEKVNGLSEGVFLPNNVKYKHGPGRWVLLAA 60
QY 61 VLIGLLVLLGIGFLVWHLQYRDVRVQKVFNGYMRITNENFVDAYENSNTFVSLASKV 120

Db 61 VLIGLLVLLGIGFLVWHLQYRDVRVQKVFNGYMRITNENFVDAYENSNTFVSLASKV 120
QY 121 KDALKLLYSYGVPLPGPYHKSASVATFSEGSVIATYWSSEFSIPQHLVSEAEVMAERVVM 180
Db 121 KDALKLLYSYGVPLPGPYHKSASVATFSEGSVIATYWSSEFSIPQHLVSEAEVMAERVVM 180
QY 181 LPPRARSLSKGFVVTSSVAFPTDSKTQVTDQNSCSFGLHARGVELMRFTTFGFPDPSYPA 240
Db 181 LPPRARSLSKGFVVTSSVAFPTDSKTQVTDQNSCSFGLHARGVELMRFTTFGFPDPSYPA 240
QY 241 HARCQWALRGDADSVLSLTFRSPDLASCDERGSGLVTVYNTLSMPHEHALVOLCGTTPPS 300
Db 241 HARCQWALRGDADSVLSLTFRSPDLASCDERGSGLVTVYNTLSMPHEHALVOLCGTTPPS 300
QY 301 YNLTFFHSQNVLLITLINTERRHPGFENFQFLPRMSSCGGRKLRKQGTNSPYPGHY 360
Db 301 YNLTFFHSQNVLLITLINTERRHPGFENFQFLPRMSSCGGRKLRKQGTNSPYPGHY 360
QY 361 PPNIDCTWNIIEVNNQHVKSFKFFYLLEPGVPAGTCTGRCIRKELRCDGWADCTDH 420
Db 361 PPNIDCTWNIIEVNNQHVKSFKFFYLLEPGVPAGTCTGRCIRKELRCDGWADCTDH 420
QY 421 NSNKITVRPHSDQSYTDTGFLAEVLSYDSSDPGCPGQFTCTGRCIRKELRCDGWADCTDH 480
Db 421 NSNKITVRPHSDQSYTDTGFLAEVLSYDSSDPGCPGQFTCTGRCIRKELRCDGWADCTDH 480
QY 481 SDELNCSADAGHOFTCKNFKCKPLFWVCDSDVNDGNSDEQSCSPAQTFRCNSGKCLSK 540
Db 481 SDELNCSADAGHOFTCKNFKCKPLFWVCDSDVNDGNSDEQSCSPAQTFRCNSGKCLSK 540
QY 541 SQQNGKDCDGDSDASCPKVNVTCTKHTYRCLNGLCLSKGNPECDGKEDCSGSDSK 600
Db 541 SQQNGKDCDGDSDASCPKVNVTCTKHTYRCLNGLCLSKGNPECDGKEDCSGSDSK 600
QY 601 DCCGLRSFRQARVVGSTDADEGEWPQVSLHALGQGHICGASLISPNMLVSAHCYID 660
Db 601 DCCGLRSFRQARVVGSTDADEGEWPQVSLHALGQGHICGASLISPNMLVSAHCYID 660
QY 661 DRGFRYSDPQTQWTAFLGLHDQSORAPGVQERLKRILSHPPFNDFPDYDIALLELEKP 720
Db 661 DRGFRYSDPQTQWTAFLGLHDQSORAPGVQERLKRILSHPPFNDFPDYDIALLELEKP 720
QY 721 AYSVMVRPCLPDASHVFPAGKAIWVTGHTQYGGTGALILQKGBIRVNTQTCENLL 780
Db 721 AYSVMVRPCLPDASHVFPAGKAIWVTGHTQYGGTGALILQKGBIRVNTQTCENLL 780
QY 781 PQOITPRMCMVGLSGVDSQCDSCGGLSSVADGRIFQAGVVSWDGCAQRNKPQVYT 840
Db 781 PQOITPRMCMVGLSGVDSQCDSCGGLSSVADGRIFQAGVVSWDGCAQRNKPQVYT 840
QY 841 RLPLFRDWIKENTGV 855
Db 841 RLPLFRDWIKENTGV 855
RESULT 14
ADB97551
ID ADI10371 standard; protein; 855 AA.
XX
AC ADI10371;
XX
DT 22-APR-2004 (first entry)
XX
DE Human cell surface protease #1.
XX
KW therapeutic agent; plasmin; protease specific antigen; PSA;
KW cell-surface protease-associated disease; cancer; ocular disease;
KW cardiovascular disease; chronic inflammatory disease; wound;
KW circulatory disorder; dermatological disorder; rheumatoid arthritis;
KW psoriasis; diabetic retinopathy; pterygium;
KW excimer laser surgery scarring; glaucoma filtering surgery scarring;
KW macular degeneration; crest syndrome; solid neoplasm; vascular tumour;
KW melanoma; Kaposi's sarcoma; human; cell surface protease.

XX Homo sapiens.
XX WO200295007-A2.
XX 28-NOV-2002.
XX 23-MAY-2002; 2002WO-US016819.
XX 23-MAY-2001; 2001US-0293267P.
XX (CORV-) CORVAS INT INC.
XX Madison EL, Semple JE, Vlasuk GP, Kemp SJ, Komandla M, Siev DV;
PI WPI; 2003-221280/21.
DR N-PSDB; ADI10370.
XX Novel conjugate useful for treating cell-surface protease-associated
PT disease, comprises a therapeutic agent and a peptidic or nucleic acid
PT substrate linked to it optionally by a peptidic linker.
XX Claim 9; SEQ ID NO 2; 581pp; English.
CC The invention comprises a conjugate that consists of a therapeutic agent
CC and a peptide substrate (optionally linked via linker). The peptide
CC substrate is proteolytically cleaved by a cell surface protease pr a
CC soluble, released or shed form of it, to liberate the therapeutic agent,
CC the conjugate of the invention is not substantially cleaved by plasmin or
CC protease specific antigen (PSA). The conjugate of the invention is useful
CC for treating a cell-surface protease-associated disease such as: cancer,
CC ocular diseases, cardiovascular diseases, chronic inflammatory diseases,
CC wounds, circulatory disorders, dermatological disorders, rheumatoid
CC arthritis, psoriasis, diabetic retinopathies, recurrence of pterygium,
CC scarring from excimer laser surgery, scarring from glaucoma filtering
CC surgery, macular degeneration, crest syndromes, solid neoplasms, vascular
CC tumours, melanoma and Kaposi's sarcoma. The present amino acid sequence
CC represents a human cell surface protease.
XX Sequence 855 AA;
Query Match 100.0%; Score 4681; DB 7; Length 855;
Best Local Similarity 100.0%; Pred. No. 2.1e-307; Mismatches 0; Gaps 0;
Matches 855; Conservative 0; Indels 0;
QY 1 MGSDRARKGGGPKDFGAGLKYNRHEKVNGLGEGVEFLPVNNVKVEKHGPGRWVLA 60
DB 1 MGSDRARKGGGPKDFGAGLKYNRHEKVNGLGEGVEFLPVNNVKVEKHGPGRWVLA 60
QY 61 VLIGLLLVLLGIGFLVWHLQYRDVRVQVFNQYMRITNENFVDAYENSNSSTEFVSLASKV 120
DB 61 VLIGLLLVLLGIGFLVWHLQYRDVRVQVFNQYMRITNENFVDAYENSNSSTEFVSLASKV 120
QY 121 KDALKLLYSGVPFLGPHYKESAVTAFSGSVIAYYMWSEFSIPOHLVEAEERVAERVVM 180
DB 121 KDALKLLYSGVPFLGPHYKESAVTAFSGSVIAYYMWSEFSIPOHLVEAEERVAERVVM 180
QY 181 LPPRARSLSKSVTVSVVAFPTDSKTQVQTNQSCSFGHLHARGVELMRTTTPGFPDPSYPA 240
DB 181 LPPRARSLSKSVTVSVVAFPTDSKTQVQTNQSCSFGHLHARGVELMRTTTPGFPDPSYPA 240
QY 241 HARCQWALRGDADSVLSTFTFRSFDLASCDERGSDLVTYNTLSPMPEHALVOLCGTTPPS 300
DB 241 HARCQWALRGDADSVLSTFTFRSFDLASCDERGSDLVTYNTLSPMPEHALVOLCGTTPPS 300
QY 301 YNLTFHSSONVLIITLINTERRHPGFEATFQPRMSSCGRLRKAQGTNSPYIPGHY 360
DB 301 YNLTFHSSONVLIITLINTERRHPGFEATFQPRMSSCGRLRKAQGTNSPYIPGHY 360
QY 361 PPNIDCTWNIEVNNQHVKVSFKFYLLPGVPAGTCKPDVVEINGEKYCGERSQFVVT 420
DB 361 PPNIDCTWNIEVNNQHVKVSFKFYLLPGVPAGTCKPDVVEINGEKYCGERSQFVVT 420

QY 421 NSNKITVRFHSDQSYTDTGFLAEYLSYDSSDPCPGQFTCRGRCIRKELRCGWDCTDH 480
DB 421 NSNKITVRFHSDQSYTDTGFLAEYLSYDSSDPCPGQFTCRGRCIRKELRCGWDCTDH 480
QY 481 SDELNCSCDAGHOFCKNKECKPLFWVCDSDVNDGDNDSDEQSCPAQTFRCNSGKCLSK 540
DB 481 SDELNCSCDAGHOFCKNKECKPLFWVCDSDVNDGDNDSDEQSCPAQTFRCNSGKCLSK 540
QY 541 SQQCNKGKDDCGDGSDEASCPKVNVTCTKHYRCLNGLCLSKGNPECDGKEDCDGSDSEK 600
DB 541 SQQCNKGKDDCGDGSDEASCPKVNVTCTKHYRCLNGLCLSKGNPECDGKEDCDGSDSEK 600
QY 601 DCDGLSFRTRQARVVGTTADGEGWPQVSLHALGQGHICGASLIISPNNMLVSAHCYID 660
DB 601 DCDGLSFRTRQARVVGTTADGEGWPQVSLHALGQGHICGASLIISPNNMLVSAHCYID 660
QY 661 DRGFRYSDPTQWTAFLGLHDQSORSAFGVQERRLRIISHPPFNDFDXYDIALLELEKP 720
DB 661 DRGFRYSDPTQWTAFLGLHDQSORSAFGVQERRLRIISHPPFNDFDXYDIALLELEKP 720
QY 721 AEYSSMVRPCLPDASHVFPAGKAIWVTGWHQTYGCTGALILQKGEIRVINTTCENLL 780
DB 721 AEYSSMVRPCLPDASHVFPAGKAIWVTGWHQTYGCTGALILQKGEIRVINTTCENLL 780
QY 781 PQQITPRMCMVGLSGGVDSCQDSDGGLSSVEADGRIFQAGVVSWGDGCAQRNKPQVYT 840
DB 781 PQQITPRMCMVGLSGGVDSCQDSDGGLSSVEADGRIFQAGVVSWGDGCAQRNKPQVYT 840
QY 841 RLPLFRDWIKENTGV 855
DB 841 RLPLFRDWIKENTGV 855
RESULT 15
ADG65326
ID ADG65326 standard; protein; 855 AA.
XX AC ADG65326;
XX DT 11-MAR-2004 (first entry)
XX DE Human MTSP1.
XX KW human; type II membrane-type serine protease 12; MTSP12; chromosome 19;
KW protease domain; MTSP12-PD1; MTSP12-PD2; MTSP12-PD3; lung; esophageal;
KW prostate; colon; ovary; cervix; breast; pancreas; cancer; tumour;
KW neoplastic condition.
XX OS Homo sapiens.
XX Key Location/Qualifiers
XX Domain 615..855
XX /label = Protease domain
XX WO2003104394-A2.
XX 18-DEC-2003.
XX 21-MAY-2003; 2003WO-US016181.
XX 21-MAY-2002; 2002US-0382851P.
XX (DEND-) DENDREON SAN DIEGO LLC.
XX Madison EL, Ong EO;
XX WPI; 2004-062325/06.
XX New type II membrane-type serine protease 12 proteins and nucleic acids,
PT useful in diagnostics, particularly for diagnosing lung, esophageal,
PT prostate, colon, ovary, cervix, breast and pancreas cancers.
XX Disclosure; SEQ ID NO 2; 207pp; English.
PS

xx	This sequence represents a human type II membrane-type serine protease 1 (MTSP1). The MTSP1 protein sequence was used in the isolation of the coding sequence of MTSP12, which has been isolated to chromosome 19. MTSP12 does not include the sequence of amino acids Arg-Lys-His-Leu-Pro-Arg-Pro-Ala ADG65347. The MTSP12 coding sequence was identified by using the protein sequence of the protease domain of MTSP1 to search the human HTGS database which produced three serine proteases, MTSP12-PD1, MTSP12-PD2 and MTSP12-PD3. EST's were identified which matched portions of MTSP12-PD1, MTSP12-PD2 and MTSP12-PD3. MTSP12 polypeptides, proteins and nucleic acids are useful in diagnostics, particularly for diagnosing lung, oesophageal, prostate, colon, ovary, cervix, breast and pancreas cancers. These are useful in immunoassays to detect, prognosis, diagnose, or monitor various conditions, diseases, and disorders affecting MTSP12 polypeptide expression, or monitor their treatment. Modulators of MTSP12 are useful for treating cancer, tumour and other neoplastic conditions.									
xx	SQ Sequence 855 AA;									
	Query Match	100.0%;	Score 4681;	DB 8;	Length 855;					
	Best Local Similarity	100.0%;	Pred. No. 2.1e-307;							
	Matches 855;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;					
Qy	1	MGSDRARKGGGKDFGAGLKNRHEKNGLEGEVEFLPVNNVKKVKKHGGGRVVLAA	60							
Db	1	MGSDRARKGGGKDFGAGLKNRHEKNGLEGEVEFLPVNNVKKVKKHGGGRVVLAA	60							
Qy	61	VLIGLLLVLLGLGFLVHLYRDVQVFNQYMTNENFVDAYENSNSTEFVSLASKV	120							
Db	61	VLIGLLLVLLGLGFLVHLYRDVQVFNQYMTNENFVDAYENSNSTEFVSLASKV	120							
Qy	121	KDALKLLYSGVDFLPGFYHKESAVTAFSEGSVIAYYWSBFSPQHLVEAEERVAERVVM	180							
Db	121	KDALKLLYSGVDFLPGFYHKESAVTAFSEGSVIAYYWSBFSPQHLVEAEERVAERVVM	180							
Qy	181	LPPRARSLKSFVTSVVAFPDTSKTVQRTQDNCSFGLHARGVELMRFTTGFDPDSPYPA	240							
Db	181	LPPRARSLKSFVTSVVAFPDTSKTVQRTQDNCSFGLHARGVELMRFTTGFDPDSPYPA	240							
Qy	241	HARCOMALRGDADSVLSLTFRSFDLASCDESGDLVTVTNLSMPEHALVQLCGTYPPS	300							
Db	241	HARCOMALRGDADSVLSLTFRSFDLASCDESGDLVTVTNLSMPEHALVQLCGTYPPS	300							
Qy	301	YNLTFHSSQNVLLITLITNERRHPGFATFFQLPRMSSCGRLRKAQGFNSPYPGHY	360							
Db	301	YNLTFHSSQNVLLITLITNERRHPGFATFFQLPRMSSCGRLRKAQGFNSPYPGHY	360							
Qy	361	PNPIDTWNIEVNNQHVKSFKFYLLPGVPAGTCKDYVEINGEKYCGERSQFVVT	420							
Db	361	PNPIDTWNIEVNNQHVKSFKFYLLPGVPAGTCKDYVEINGEKYCGERSQFVVT	420							
Qy	421	NSNKITVRFHSDQSYTDGFLAEVLSYSDSDPCQFTCTGRCIRKELCDGWADCTDH	480							
Db	421	NSNKITVRFHSDQSYTDGFLAEVLSYSDSDPCQFTCTGRCIRKELCDGWADCTDH	480							
Qy	481	SDELNCSADAGHQFTCKNFKCKPLFWVCDSDNDSDGSCPCPAQTFRCNSGKCLSK	540							
Db	481	SDELNCSADAGHQFTCKNFKCKPLFWVCDSDNDSDGSCPCPAQTFRCNSGKCLSK	540							
Qy	541	SQCNKGKDDCGGSDSEACPKVNVVCTTKHYRCLNGLCLSKGNPECDGKEDCSGDEX	600							
Db	541	SQCNKGKDDCGGSDSEACPKVNVVCTTKHYRCLNGLCLSKGNPECDGKEDCSGDEX	600							
Qy	601	DCDCGLRSFTQARVVGTTDADEGEWPMQVSLHALGQGHICGASLIISPNWLVSAAHCYID	660							
Db	601	DCDCGLRSFTQARVVGTTDADEGEWPMQVSLHALGQGHICGASLIISPNWLVSAAHCYID	660							
Qy	661	DRGFYSDPQTQWTFALGLHDQSQRSAQVQERRLKRIISHPFFNDFTFDYDIALLELEKP	720							
Db	661	DRGFYSDPQTQWTFALGLHDQSQRSAQVQERRLKRIISHPFFNDFTFDYDIALLELEKP	720							
Qy	721	AEYSSMVRPICLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGIRVINQTTCCENLL	780							

Search completed: September 23, 2005, 12:55:36
Job time : 88 secs

Db	721	AEYSSMVRPICLPDASHVFPAGKAIWVTGWGHTQYGGTGALILQKGIRVINQTTCCENLL	780
Qy	781	PQOITPRMNCVGFSLSGGYDSCQDSCGGPLSSVEADGRIFQAGVVSWGDCGAQRNKPQVYT	840
Db	781	PQOITPRMNCVGFSLSGGYDSCQDSCGGPLSSVEADGRIFQAGVVSWGDCGAQRNKPQVYT	840
Qy	841	RLPLFRDWIKENTGV	855
Db	841	RLPLFRDWIKENTGV	855